# TECHNOLOGY BASED AID FOR THE VISUALLY IMPAIRED

by

Eoghan Martin

A dissertation submitted to The University of Dublin in conformity with the requirements for the degree of Master in Computer Engineering.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Trinity College Dublin

May, 2016

# Declaration

I, Eoghan Martin, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

May 19, 2016

Eoghan Martin

# Abstract

This dissertation presents a system that can assist a person with a visual impairment in both navigation and mobility. To date, a reliable solution has not been put forward to replace the legacy devices currently used in mobilizing and navigating on a daily basis for people with a visual impairment.

This report first examines the problem at hand and the motivation behind addressing it, then explores relative current technologies and research in the assistive technologies industry and finally proposes a system design and implementation for the assistance of visually impaired people.

The system proposed in this project includes the use of Google's Project Tango and Cloud Vision APIs as well as the use of a smartphone for haptic relay of information to the user. Implemented, is a home-made stereoscopic camera system that uses a laptop to relay environment information to a user via haptic feedback on a smartphone. A series of experiments are run to evaluate the implemented system and future work for the project is documented.

# Acknowledgments

I would like to thank Dr. Donal O'Mahony for his help and guidance on this project. A special thanks to the NCBI for their feedback and interaction. I would also like to thank my family and friends for their support and proof readings.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Motivation

There are an estimated two hundred and forty-six million visually impaired people worldwide [14]. There are two main mobility aids used by people who are visually impaired - the long cane and guide dogs. These devices do not exploit technologies of modern society to their advantage.

Leveraging the potential of modern technologies, this report explores how much they can be used to replace the current legacy devices used by visually impaired people for mobility and ultimately improve the quality of life for these people worldwide.

### Problem

Visually impaired people worldwide are using outdated devices for mobility and navigation which limits their quality of life.

**Hypothesis**

Modern technologies can be used to improve the quality of life for visually impaired people.

## 1.2 Background

### 1.2.1 Visually Impaired

An impairment can be looked at from two main approaches [15] - the medical and social models. The medical model is defined by the WHO [1] [16] as:

> *"Any loss or abnormality of psychological, physical or anatomical structure or function."*

The social model is defined by the DPI [2] [17] as:

> *"The functional limitation caused by physical, sensory or mental impairments."*

A good illustration of the difference between these two outlooks is put forward in [15] with the example of reading a book. The medical approach to an impairment defines that a person is visually impaired if they cannot read a book with standard text sizes. The social approach puts forward the hypothesis that a person is visually impaired because only some books are available with a large font. If books were more accessible, the social approach would then define many people to cease being impaired. It can therefore be concluded that

---

[1] WHO - World Health Organization

[2] DPI - Disabled Peoples International

if the world we live in becomes more accessible to people with weaker vision, the number of ”visually impaired” people will decrease - resulting in a higher quality of life.

According to a fact sheet put forward by the WHO [14], there are an estimated two hundred and forty-six million visually impaired people in the world. This fact sheet also provides an insight to the demographic of visually impaired across the world. About ninety percent of visually impaired people are from low-income backgrounds in the first world. As well as this, eighty-two percent of people living with blindness are aged fifty and over. When designing a visual aid, it is important to keep these attributes in mind. On top of all of these numbers, it is estimated that by the year 2020, the number of visually impaired people will double [18].

The main issues for accessibility of a visually impaired or blind person are split into six activities - mobility, communications and access to information, cognitive activities, daily living, education and employment, and recreational activities [19]. The barriers for each activity are shown in Table [1.1].

**Table 1.1:** Activities and their main accessibility barriers.

| **Activity** | **Barrier** |
|---|---|
| Mobility | Obstacle avoidance |
| | Navigation & orientation |
| | Access to environment |
| Communication and information access | Access to print media |
| | Computer and internet access |

**Table 1.1** . . . continued

| Activity | Barrier |
|---|---|
| Cognitive activities | Analyzing information<br>Logical and creative thinking |
| Daily living | Personal care<br>Environmental control |

The mobility activity will be focused on with the goal of improving the quality of life for people with a visual impairment or blindness.

When moving around on a daily basis there are a few key obstacles that must be navigated about to ensure safe travel. It is these fundamental obstacles that should be designed for in the development of a visual aid for the visually impaired. The obstacles are listed in Table 1.2.

**Table 1.2:** Key areas of obstacle avoidance on a daily basis.

| **Area** | **Obstacle** |
|---|---|
| Leg Height | Litter Bins<br>Path Curb Stones<br>Path Drop Off and Stairs<br>Slopes and Inclines |
| Head Height | Tree Branches<br>Archways |
| Side Obstacles | Scaffolding or a Wall<br>Doorway<br>Narrowing Path |
| Destination Reaching | Pedestrian Crossing<br>Door |

**Table 1.2** . . . continued

| Area | Obstacle |
|---|---|
| | Waste Bin |
| | End of Pavement |
| | Street Furniture |

### 1.2.2 Assistive Technology

Assistive technologies are becoming a large and well funded industry. ATIA [20] [3] is an industry standard establishing organization that was founded in 1998 for the global assistive technology market. This organization produces an academic journal entitled "The Assistive Technology Outcomes and Benefits Journal" and runs an annual conference where leaders in this field can present their work and others with common goals in this industry. ATIA has helped standardize the market for assistive technologies and progress this area as a whole. Since its founding, there have been many other conferences for assistive technologies [21], [22] and [23]. As well as conferences, there are global awards for achievements in the area of assistive technologies - [24] and [25].

Since the founding of these bodies, conferences and awards the industry has grown to be large as illustrated by a report published by BCC Research [26];

> *"The U.S. market for assistive technologies is projected to grow from \$40.6 billion (including eyeglasses and contact lenses) in 2014 to \$43.1 billion in 2015 and \$58.3 billion in 2020, with a compound annual growth rate (CAGR) of 6.2% between 2015 and 2020."*

---

[3] Assistive Technology Industry Association

### 1.2.3 Wearable Technology

Wearable devices are articles of clothing or accessories that incorporate electronic technologies and computing power. This type of device is on an upward trend and is beginning to become a part of our day to day lives. PWC [4] worked with BAV Consulting [5] to conduct a survey of a very large and diverse sample population and used the collected data to publish a report on the future of wearables [27]. According to this report:

> *"Our research shows that there is a wearable future around the corner, its more immediate than we think and it can dramatically reshape the way we live and do business."*

According to research conducted by the IDC [6] [28], shipments of smart wearables have over doubled since 2014 and will continue to grow by nearly 400% per year thereafter.

Furthermore, this IDC research concluded that by 2019 there will be one hundred and fifty million wearable devices worldwide and one hundred and twenty million of those devices will be wrist-wear - this is a growth from twenty million wrist-wear devices in 2014. On top of the IDC research, according to a study conducted by WiForce [29] [7], the global revenues for smart watches are growing and will continue to grow until 2020. This will, in turn, reduce the cost of these devices and improve the technology available to this type of wrist-wear.

---

[4] PWC - PricewaterhouseCoopers, LLP (accounting/consulting firm)

[5] BAV Consulting - a global leader in research

[6] IDC - International Data Corporation

[7] Figure 20, Page 49

As a result of this upward trend, this report will focus on wearable technologies in the development of a visual aid.

## 1.3 Project Goals

This project seeks to demonstrate how technology can be used to improve the quality of life for people with a visual impairment.

It should act as a platform to gain an understanding of the capabilities that modern technologies will possess to assist people with impairments in the future.

The key research objectives which have been put forward for this dissertation are:

- Assess the main needs for a person with a visual impairment.
- Assess the current options on the smart wearables market and within research.
- Gain an understanding of the capabilities that sensors currently have.
- Develop the functionality for a technical solution.
- Assess the possibilities in the future for this technology.

## 1.4 Outline

**Assistive Technologies Research**

This chapter begins with an overview of a focus group held with the NCBI. It then investigates assistive technologies via state of the art and literature review sections.

**Design**

This chapter illustrates the early design for this project as well as its pros and cons. It then goes on to detail and investigate final design proposed for this project.

**Implementation**

The implementation is laid out in its core components; the Python application divided into its fundamental sections followed by the Android application and its main sections.

**Evaluation**

The evaluation provides an overview of the technical implementation for this project. It details experiments that were run and limitations of the system.

**Conclusion**

The final chapter draws conclusions from the evaluation section. It also acts as an overview of the goals set out at the start and defines future goals for this project.

# Chapter 2

# Assistive Technologies Research

Current tools of aid available to people with a visual impairment are investigated, specifically in the wearable technology sector. The main needs of people with visual impairments and their biggest obstacles on a day-to-day basis are also explored.

According to a study conducted by the NCBI [30] [1] in association with The Discipline of Occupational Therapy, Faculty of Health Sciences, University of Dublin, Trinity College 25.4% of young blind people (from a survey of 564 blind people in Ireland) do not leave their local area without a sighted guide. This is mainly down to - as mentioned in Section 1.2.1 - a lack of mobility accessibility.

The two main tools for obstacle avoidance today are the long cane and guide dogs. According to the study carried out by NCBI [30], most people with a visual impairment from both the young and elderly bracket use their residual

---

[1] NCBI - National Council for the Blind Ireland

vision and/or a sighted guide to navigate their way around on a day-to-day basis. Only about 28% of the visually impaired people from this study use an aid for mobility - the long cane being the most popular tool in aiding mobility at 18%.

The long cane can only feedback information about the area up to one stride ahead of the user. This can sometimes be increased by tapping the cane and calculating a frame of reference by analyzing the echos of the sound. The long cane is limited in its accuracy and resolution and can not be used effectively for navigation purposes - for calculating the best route to a location.

## 2.1 NCBI Investigation

A two-hour interview with some visually impaired members of the organization [31] was carried out.

The findings are as follows:

- Some of the members do not like to use a visual aid that is very visible.
- Assistive technologies are very expensive with a simple text reader costing $99 [32].
- Reading small text is an issue, especially when shopping and buying food.
- Not understanding where a path may end along a roadside - This was the most prominent issue put forward in this meeting.

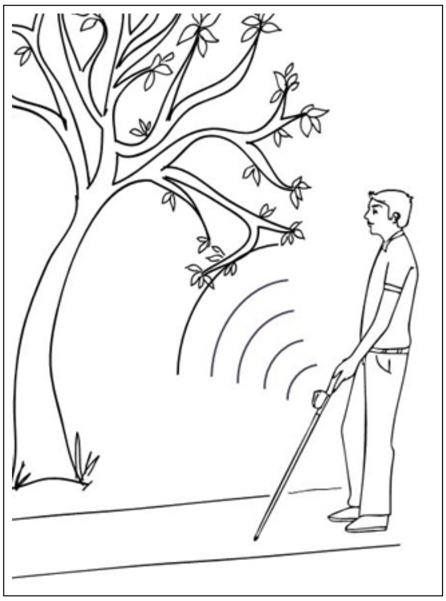


**Figure 2.1:** SmartCane functionality [1].

## 2.2 State of the Art

1. **Laser Cane**

   The main two types of laser canes are the SmartCane [33] and UltraCane [34]. These devices use ultrasonics to help detect obstacles in the way of the user. They cost less than 100 euros and are a helpful addition to the traditional long cane. This device works by giving a haptic feedback to the user using vibration patterns to inform them when obstacles are approaching. The ultrasonic sensors are embedded in the handle and are directed to the front of the user as shown in Figure 2.1. This type of device proposes the idea of customizing and improving a current aid device using modern technologies.

2. **BlindSquare and Seeing Assistant**

BlindSquare [35] and Seeing Assistant [36] use GPS to improve navigation for the visually impaired. These devices gather information from Google Maps, Yelp and Foursquare about the location in which the user is situated. The software works by creating an audio narration to the user about their location. This type of technology has been proven to help people find their bearings in locations that have a lot of information on Google Maps, Yelp and Foursquare but cannot provide the user with much information outside these locations.

3. **Sentiri**

Chaotic Moon's Sentiri [37] is a product that was designed to assist the visually impaired. The device is made up of a 360-degree headband that uses IR sensors to detect obstacles and uses haptic feedback on the head to inform the user where the closest obstacles are. It is a difficult technology to use outdoors as a result of the use of IR sensors. IR sensors send and receive pulses of light at a wavelength range of 850nm +/-70nm. The sun's rays may also modulate at these frequencies causing an interference [2].

This device can only detect obstacles accurately that are close to the upper body. This proved to be a design flaw and must be considered when designing a visual aid.

4. **Wayfindr**

---

[2] This will be explored further in Section 2.4

Wayfindr [38] uses audio as an aid to visually impaired people to help navigate through busy environments. Wayfindr beacons are installed in busy and difficult to navigate areas like train stations. These beacons relay the information via Bluetooth to Wayfindr users in the vicinity. This method of aid has proven to help people in the Pimlico Underground [39], however, it is not scalable.

5. **OrCam**

   OrCam [2] uses vision technology to let the user know what objects or information is in the scene around them. A camera is fixed to a pair of glasses. When the user points to an object in view of the camera as shown in Figure 2.2, the OrCam software will use OCR (Optical Character Recognition) and object recognition to generate relative information about the area being pointed to. Audio is used to relay this information to the user. This is not majorly helpful for ease of mobility.

6. **Project RAY**

   Project RAY [3] have developed a variety of products targeted at the visually impaired. The product base of Project RAY is made up of both software and hardware. The software includes a mobile application tailored to people with debilitated vision. This smartphone application provides an easy to use interface to allow visually impaired people to use all of the main features of a phone easily. It acts as a skin on top of the phone's

**Figure 2.2:** OrCam usage [2].

operating system and provides easy access to functionality like calling, messaging and using maps. An example of the home screen is shown in Figure 2.3.

7. **Cities Unlocked**

   Cities Unlocked [40] is an initiative that brings together Microsoft, Future Cities Catapult, and Guide Dogs to make cities more accessible. For this project, Microsoft has developed a prototype of a product that uses audio to give the user an image of their environment. It is referred to by Microsoft as a 3D soundscape. The software in this device uses Bing maps and Bluetooth beacons on lamp posts to gather information about the area around a user and feed it back to them using bone-conduction

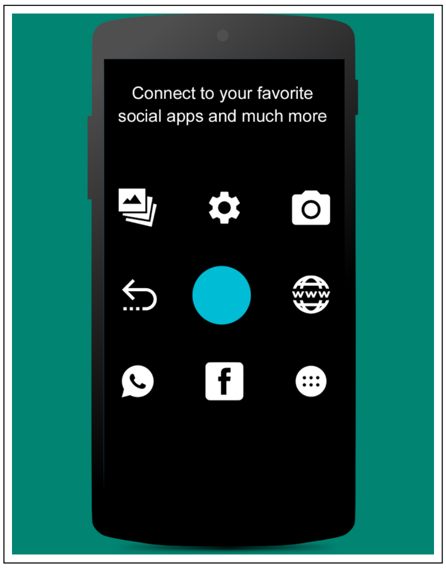


**Figure 2.3:** Project RAY home interface [3].

techniques.

Working with a company called AfterShokz [4], the device sends vibrations to the upper jawbone to communicate with the user. According to AfterShokz,

> *"Transducers guide mini vibrations through the cheekbones to the inner ears, delivering sound without plugging or covering them."*

This by-passes the ear drum and allows the user to hear audio as ambient sound. This project demonstrates how significant sound is as a sense and how it can be leveraged to make environments more accessible to the visually impaired.

8. **Horus Technology**

Horus [5] is a head wearable that hooks over the ears of the user. Horus

**Figure 2.4:** AfterShokz headphones are placed on the jawbone just below the ear [4].

uses cameras to detect objects in a scene as well as faces and text [3]. It then feeds the information from the scene back to the user via sound. This is a step further than the Cities Unlocked project described in Section 7 in that Horus uses cameras and computer vision, as well as 3D sound to relay information to the user via bone-conduction. The technology setup can be seen in Figure 2.5. The method of sound communication used in this technology allows the user to hold a conversation or listen to music while still hearing the 3D sound relay of information.

9. **Lechal**

Lechal [41] is a smart device referred to as an interactive haptic feedback footwear. This is smart device in the insole of a shoe that can connect to a smartphone via Bluetooth. Vibrations to the feet are used to inform the user which direction to walk. This works effectively for macro navigations

---

[3] The computer vision techniques used will be investigated in Section 2.4.2

**Figure 2.5:** Horus Technology headset contains a camera with bone conductors to be placed on the jawbone, just below the ear [5].

but has not been implemented for mobilizing a user around obstacles in the immediate environment. By using short bursts of a variety of vibrations to the smart insole, a device may be able to navigate a user around close proximity obstacles.

10. **Project Tango**

    Google has been developing a super sensing device in a project named Project Tango [42] for over ten years. This project involves building a smart device that is highly aware of its surroundings. With this heightened perception, the device is perfect for the incorporation of augmented reality. Project Tango involves the development of a smart device [4] containing highly customized hardware and software to perform three primary functions - motion tracking, depth perception and area learning. The primary sensors included in the device are a motion tracking camera, 3D depth sensing camera, accelerometer, ambient light, barometer,

---

[4] Originally a smartphone with added sensors but now a larger tablet device.

compass, GPS, and gyroscope [5]. The device senses depth using two cameras in stereo as well as an IR sensor. On top of this, the device has a high-end NVIDIA graphics processor included and a Movidius [6] VPU (Vision Processing Unit) for processing specific computer vision tasks more efficiently. This device is powerful enough to capture an entire room in three dimensions and convert it into a computerized mesh just from walking around the room. The device tracks where the user moves and can add additional information to areas that have been previously scanned to improve its readings. The Project Tango device can accurately measure objects in an image and calibrate them so as to convert virtual measurements to measurements that can be easily quantified by humans,like the metric system.

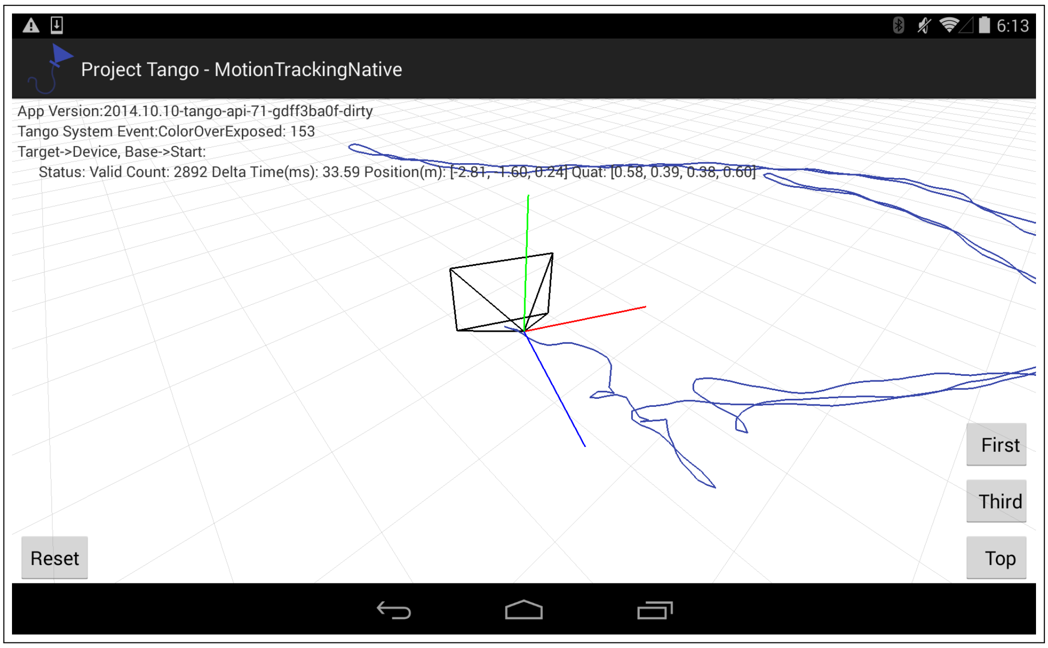
## 2.3 Biomedical Developments

Research into biology and its use when paired with computer science is growing rapidly. This research crosses paths with many of the technology aspects covered in the previous sections. Areas like the software behind computer vision are being reevaluated by investigations in biology. A 2014 paper [43] explores the potential for biological systems like neural networks to play a role in improving computational ability in fields like machine learning and computer

---

[5] These sensors will be explored in Section 2.4

[6] Movidius have been working on optimizing artificial intelligence and computer vision based computation for over nine years.



```
App Version:2014.10.10-tango-api-71-gdff3ba0f-dirty
Tango System Event:ColorOverExposed: 153
Target->Device, Base->Start:
   Status: Valid Count: 2892 Delta Time(ms): 33.59 Position(m): [-2.81, -1.60, 0.24] Quat: [0.58, 0.39, 0.38, 0.60]
```

**Figure 2.6:** Motion tracking in the Project Tango device [6].

vision. The conclusions from this paper, however, still involve looking to the future for implementation of these systems in computer science.

1. **Argus II [44]**

   This system is made up of a retinal implant that uses electronic stimulation of the retina to communicate with the user. A series of electrical stimulations are used to communicate visual perception to a blind person. The Argus II implant is targeted at people who have lost their sight due to retinitis pigmentosa [7]. This system is an aid but allows the user to see and perceive their environment rather than providing assistance through the use of a computer. Over time, it is possible that this method of feedback to a user may be incorporated with wearable sensors to improve a

---

[7] Retinitis pigmentosa is an eye disease that causes a loss of vision as a result of damage to the retina.

user's sight even further.

2. **Pixium Vision [45]**

   Pixium Vision is a company providing a similar solution to Argus II. Pixium Vision are developing a retinal implant attached to the surface of the retina which uses electrical pulses to stimulate the nerve cells in the retina. A smart pair of glasses has also been developed which can communicate with the retinal implant and feed back information to the user if necessary as well as run computations. Some of the correction algorithms used in these glasses are similar to the stereophonic algorithms explored in Section 2.4.2.

3. **Google Intra-ocular Device Patent [46]**

   This patent implies that Google plans to go into the retina implant sector. Google have also patented a contact lens [47] that has an embedded camera which may accurately measure health parameters like glucose levels and could assist the visually impaired and blind. Samsung has patented a similar device. The use of some of the technology in Project Tango may eventually be embedded into a contact lens or a retinal implant.

## 2.4 Literature Review

This section is an exploration of research that has been carried out in the assistive technologies sector focusing specifically on the use of wearables as well as a selection of other promising technologies.

[48] documents the potential for electronics and technology to play a significant role in aiding accessibility for the blind. This 1982 paper outlines some of the wearable ideas put forward by Brabyn like spectacles with ultrasonics, a sonar system with a tactile display that's strapped to the torso and signs that transmit information to a hand-held speaker device. This report investigates how research has transformed since the early papers in the 1990's.

Throughout the 1990's there were developments in the assistive technologies space for the visually impaired but these technologies could not be developed unobtrusively. Discreetly effective technology is critical in this area as many users do not want to be seen using assistive aids. As a result of the scale of computers and the amount of computing power required, wearables were not achievable during these times.

In 1992 an image-to-sound converter [49] had the goal of representing each pixel and alternatively, set of pixels, with a different audio tone. With the lack of computer vision technologies at this time, it was an innovative way to represent an environment to a person using a small amount of computation power.

In 1995, [50] describes the development and implementation of robotic mobility aid which senses the environment to ensure that the user is safe and can avoid obstacles. This robotic device uses IR sensors as well as sonar technology to detect when an obstacle is in the way and direct the user around it. The technology used in developing this device is now very small and can be worn by a user - as opposed to the 1995 computing capabilities which were mounted on a robotic trolley. This project proved that sensors have the potential to guide a visually impaired user through an unknown environment.

Until the early 2000's, assistive technology research consisted mostly of IR sensors and sonar to understand the environment around people and guide them to safely avoid obstacles [8]. In the early twenty-first century came the inception of computer vision. This new type of technology opened up a whole new realm for the development of effective assistive technologies.

Throughout the early 2000's, development in the computer vision area became mainstream. Technologies were built like the Xbox Kinect and the Wii remote implementation. The Xbox Kinect and Wii remote use computer vision techniques to allow users to gesturally control gaming systems. These technologies use IR sensors and basic concepts of vision to understand how a user is interacting with their environment.

### 2.4.1 Sensors

#### SONAR AND ULTRASONICS

Sonar is the use of sound to echolocate objects in the distance. This technology was first introduced after the sinking of the Titanic [51] and is still used today in boats as well as many other applications.

---

[8] This is evident here [50].

**Figure 2.7:** Sonar glove design [7].

Sound travels at a uniform speed. Therefore, when it is reflected from a surface, the distance at which that surface is from the transmitter of the signal can be calculated. There are two types of sonar, active and passive sonar. Passive sonar only collects pulses and it is used for hearing whether any sonar transmitters are present in the vicinity. This is not of much use for a visual aid. Active sonar transmits and receives pulses very rapidly. It consists of a transmitter as well as a detector. The time between the sending and receiving of a pulse determines the distance at which the reflected surface is away.

[7] proposed a glove wearable that leverages the power of sonar to understand the environment around the user. The concept works similar to a long cane. An Arduino Pro Mini is used to analyze the received echo pattern. If there is an object in the immediate direction in which the glove is pointing, a buzzer will sound [2.7].

The most attractive features for sonar is that it works well both indoors and outdoors and can detect distances that are not just relative to other objects in a scene but measurements that are calibrated to the metric system. Sonar

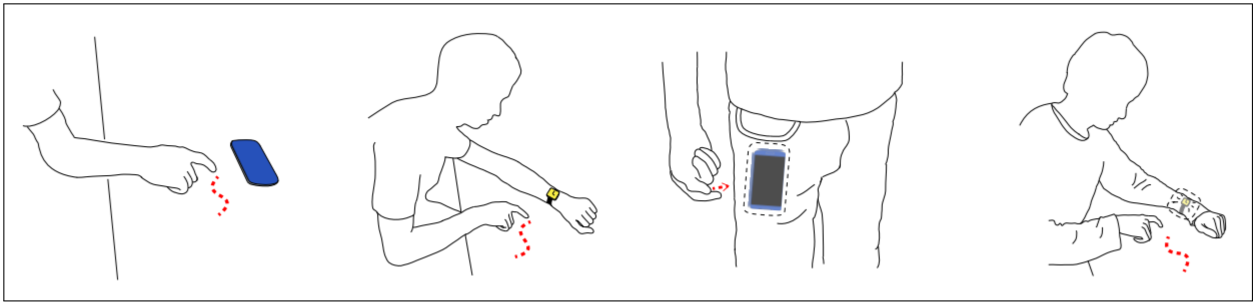


**Figure 2.8:** System presented in [8] by fingerIO.

does not pick up enough information to compete with the use of cameras and computer vision. Instead, sonar may be more useful for the user to interact with their smart devices. A paper presented in May of 2016 [8] gives an insight into the potential for a user to communicate with their smart devices without having to use the small and detailed touch screen. [8] is the basis for the creation of a software company called fingerIO and presents an effective way to track finger movements to communicate with a smart device. This may, for example, be used to change the volume on a smart watch or direct a camera on a smart device to read text in a specified direction relative to the user's position. Figure 2.8 shows the system presented in this paper. This figure also shows how the technology can be used from within the user's pocket.

## IR SENSORS

IR (infrared) sensors measure the amount of light emitted by objects in a scene. They are made up of a transmitter and a receiver. The stronger the light rays on the receiver side are, the higher the reading from the sensor will be 2.9]. These sensors can be used to detect movement in a scene and even

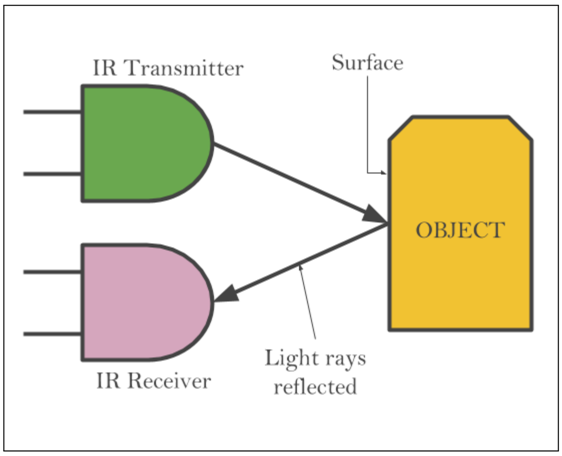


**Figure 2.9:** IR sensor functionality.

sense when objects are approaching. Some of the most popular applications for IR sensors are in the Kinect and Wii game controlling systems. The technology in these systems provide for the foundations of a technology based visual aid. The Kinect uses a combination of IR sensors and cameras, but the Wii remote is a purely IR based technology.

The Kinect uses both an IR camera and a stereo camera to build a three-dimensional representation of a scene. In Figure 2.10 both the IR camera and stereo camera can be seen. The functionality of this stereo camera will be explored in Section 2.4.2. The IR camera was developed by Prime Sense Ltd [52] and projects an IR pattern onto the scene in front of the camera. This pattern is then captured and processed to reconstruct a three-dimensional map of the object. The parts of the reflected pattern that are smaller, are further away in the scene.

Johnny Chung Lee [9] explored how an IR camera can be exploited to track

---

[9] Technical Program Lead for the Special Projects Team in Google where Project Tango is

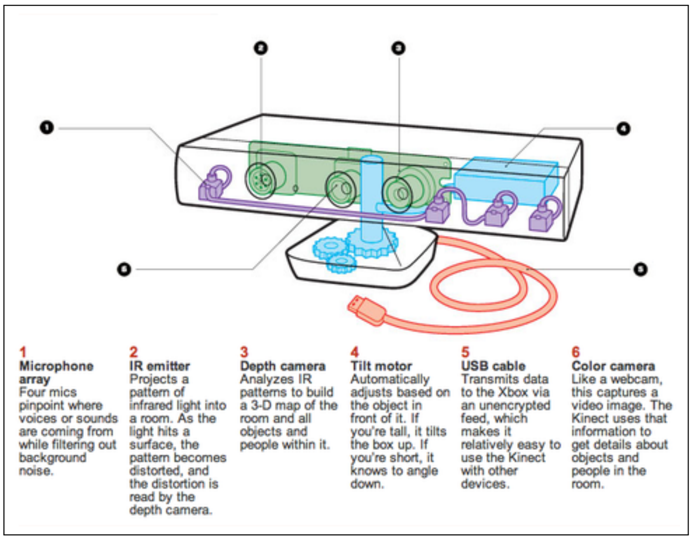


**Figure 2.10:** Kinect technology setup [9].

fingers in a scene while doing his Ph.D. in Carnegie Mellon [53]. This exploration of IR capabilities follows the fingers of a person quite accurately by applying reflective tape to the tip of the finger. This tape makes the finger more reflective and hence more vibrant to an IR sensor. Applied to assistive technologies, this may be very helpful in directing a camera on the direction to point.

Although IR cameras are very accessible as they are on most smartphones today, they are not majorly effective when used in naturally lit areas. This is as a result of the range of light frequencies provided from the sun. Sun light rays can often modulate at the same frequency as the IR sensor's (IR sensors have a wavelength range of 850nm +/- 70nm) causing interference on the receiver side of the device.

## ACCELEROMETER

---

developed.

An accelerometer is an instrument that measures the acceleration of an object in a particular direction. This acceleration is measured within a relative coordinate system and can be mapped to the global coordinate system using a compass. The accelerometer is a very suitable instrument for measuring the orientation of a device. An important functionality of an accelerometer is to detect when something is moving. This may, for instance, be important when deciding the time at which to activate a visual aid system and when to go to a power saving sleep mode. A technology based visual aid device may require large amount of power to compute vision based tasks, hence, power saving mechanisms may be important. This is also one of the key technologies in gesture recognition.

[54] funded by the VENTURI project presents a practical use of the accelerometer for the visually impaired. This paper provides an alternative to the touch interaction associated with smartphones. Specifically, this article describes a system built to use gestures to inform a smartphone about specific vision tasks to run, like for instance logo recognition. When the user moves their arm in a predefined gesture, with a smartwatch, a corresponding task can be run on their smartphone.

Another use of the accelerometer in aiding the visually impaired is presented in [55]. This article presents a method to utilize the accelerometer in a smartphone to improve the input quality of an image from the camera. The accelerometer vector is used to translate the pixels of a camera frame using the average accelerometer vector from a series of images corresponding to a series of points in time. The use of this methodology may held stabilize and reduce motion blur effects when a camera is mounted on a wearable device like a belt in aiding a visually impaired person.

**CAMERA**

Cameras are the most widely used tool in the technology-based visual aid sector. Cameras have significantly reduced in both price and size over the last few years while also increasing in quality and capability. A variety of vision-based functionality can be implemented using a camera as will be explored in the following subsections.

### 2.4.2 Computer Vision Techniques

Computer vision takes on an algorithmic approach to analyzing and extracting information from images. These algorithms are becoming more and more accurate; however, they require quite a lot of computation power.

There has been a lot of research on the use of computer vision to assist the visually impaired. There have been a lot of innovative solutions provided that involve embedding cameras in a variety of wearables. Some examples of this are the use of the Xbox Kinect mounted to the torso and around the waist of users. Some papers that presented these ideas are [10], [11] and [56].

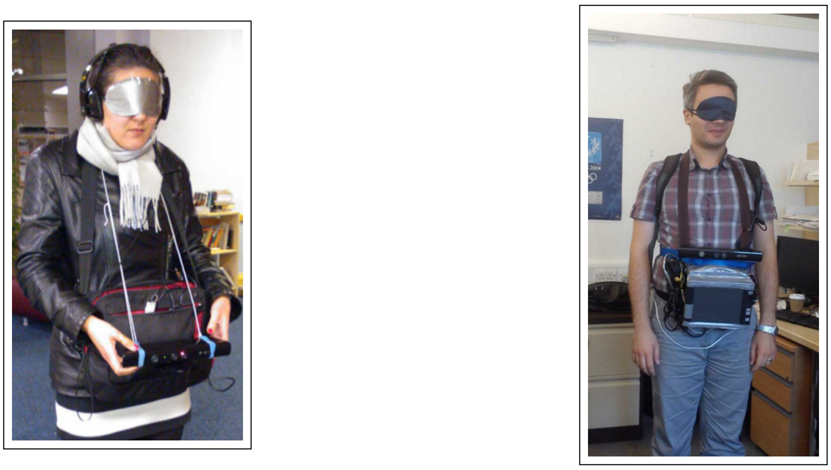


**Figure 2.11:** Mounted Kinect devices from Fig. 1 in [10] and Fig. 8 in [11]

A research team within Google[10] began to investigate the potential for a device that can read a scene and understand it in the way that the Kinect does. The project was entitled Project Tango which is explored in Section 2.2.

Computer vision is about translating pixels into abstract mathematical concepts. It is computer vision (mixed with IR sensing) that allows people to use the Kinect as an assistance tool for the visually impaired. It is also computer vision that allows speed cameras to function and eventually, it will be computer vision that allows autonomous cars to function.

### DEPTH SENSING

Depth sensing can be done using both IR sensors and sonar. These methods both have drawbacks and cannot provide the detail that camera depth sensing can provide. There are two main ways to perceive depth using a camera - using depth-of-field and a single eye, and using a stereo camera setup with multiple

---

[10] The Special Projects Team led by Johnny Chung Lee.

eyes.

Depth sensing is traditionally done in an SLR camera using depth-of-field. The depth-of-field is the distance between the nearest and farthest objects in a scene that appear with adequate sharpness in an image. This distance can be changed by varying a few attributes in the camera. For instance, by varying the aperture [11], focal length [12] and lens distance to the focus point [13], the depth of field changes [Fig. 2.12]. This technology is implemented in the Lytro [14] camera devices [58] and has been labeled light-field cameras. Ren Ng developed the technology behind Lytro during his Ph.D. studies at Stanford University [59]. The idea of using many lenses and aperture sizes to extract three-dimensional information from a scene has been explored since Gabriel Lippmann [15] proposed the idea of integral photography. In 1996, [60] brought three-dimensional capture from a single eye camera to the fore [16]. The technology was further developed by a popular dissertation presented in 2011 [61]. This type of technology is hindered by its size, as well as it's price. With a smaller sized light-field camera comes a reduced resolution. This is the foremost reason for these devices not appearing more frequently on mainstream smart devices. Some companies have taken to embedding a second, lower resolution camera on their devices so as to supplement depth perception taking a

---

[11] This is a gap behind that lens that allows light through to reach the film or sensor.

[12] This is the distance from the lens to the film.

[13] This is the point at which the majority of rays from the lens meet.

[14] Lytro's main competitor is Raytix [57]. These are currently the leaders in this field.

[15] A 1908 Nobel Laureate.

[16] This paper is the 4th most frequently cited paper in the computer graphics literature.

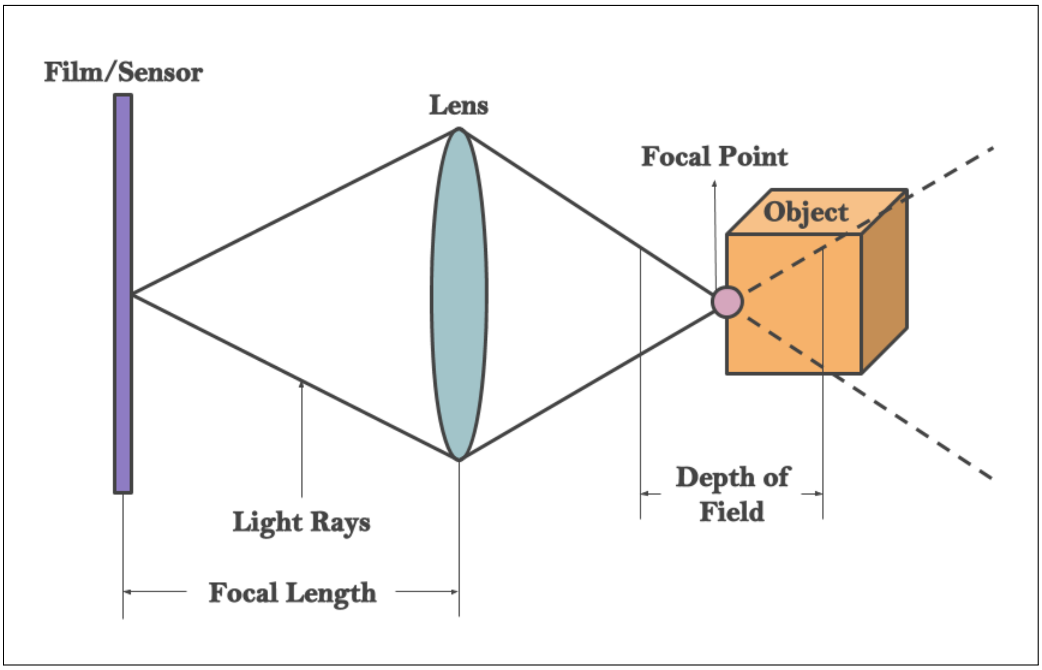


**Figure 2.12:** Parameters that are manipulated to generate 3D images.

light-field approach [17].

Although light-field cameras look promising for the future, the most popular method of depth perception using cameras involves the use of a stereo camera. This type of vision has been studied quite extensively over the past 25 years but only has a few commercial systems available today.

A stereo camera contains two image sensors and simulates binocular vision in the same way that humans see. Stereoscopic vision involves the input of two image frames and from the offset of the pixels, a three-dimensional depth plane can be calculated. The further apart that the camera sensors are, the more three-dimensional information that can be calculated - this also results in more noise so a calibrated optimum distance should be used. To be as accurate

---

[17] Project Tango uses a lower resolution depth perception camera.

**Figure 2.13:** Left and right stereo images combined to produce a disparity map. Images were produced for the stereo vision research of Daniel Scharstein, Richard Szeliski and Heiko Hirschmller [12]

as possible, both camera sensors should have the same focal length. The pixel difference in the two images is known as the disparity and using this disparity a gray-scale image can be generated which depicts the different depths of the corresponding pixels in the images. At the most basic level, this disparity is used as shown in Fig. 2.14. This figure shows how as objects in a scene get further away from the cameras, the disparity in the corresponding areas of the images will decrease. These disparity values can then be used to determine a depth point for each pixel in the scene, encoding the third dimension into the image. This third dimension can be used to extract important information for mobilizing about an environment with a visual impairment.

To use stereoscopic vision in its most efficient manner, the intrinsic parameters of each camera must be known. The captured frames must also be perfectly horizontally offset - the image planes must be coplanar with the ability to distinguish epipolar lines. These straight lines are made up of corresponding pixels from each image. Using the intrinsic parameters, measuring the image point on one camera can be used to find the corresponding point on the other.

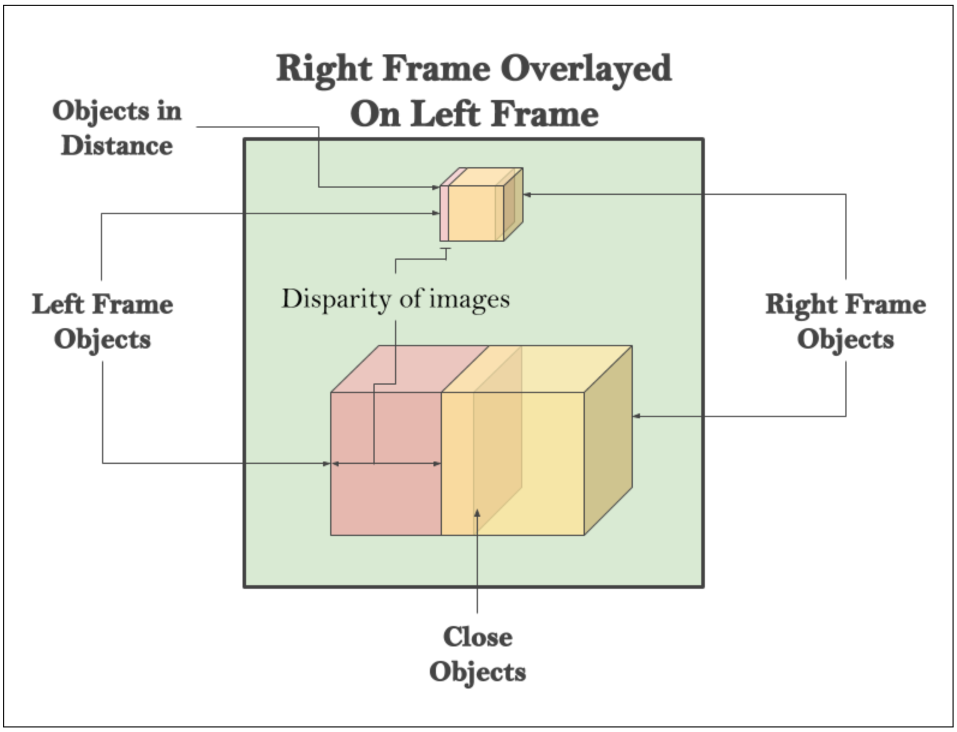


**Figure 2.14:** Disparity size with distance.

This is proven by considering a standard rig for accurate stereo cameras and similar triangles. The equations in Fig. 2.15 describe the relationship between the point P and the disparity (d) between the left and right camera frames. The final equation here describes how, as the objects in the scene get further away from the cameras, the disparity (d) decreases. This is also evident in the diagrams in Figure 2.14 and Figure 2.17.

Multi-View Stereo (MVS) algorithms which require more computational power than pure stereo vision but use only one camera [62]. This type of algorithm is used to render three-dimensional images from a single camera. The underlying principles of this technology are also employed in the utilization of a stereo camera that does not have perfectly horizontally aligned and identical

$$X_A = f.\frac{X_{PL}}{Z_{PL}} \qquad X_B = f.\frac{X_{PR}}{Z_{PR}}$$

$$d = X_B - X_A \qquad Z_{PL} = Z_{PR}$$

$$X_B = f.\frac{X_{PL} + b}{Z_{PL}} \qquad Z_{PL} = f.\frac{b}{d}$$

**Figure 2.15:** As d increases, Z decreases. Therefore, there's an increase in disparity with closer objects in a scene. These equations correspond with Figure 2.16.

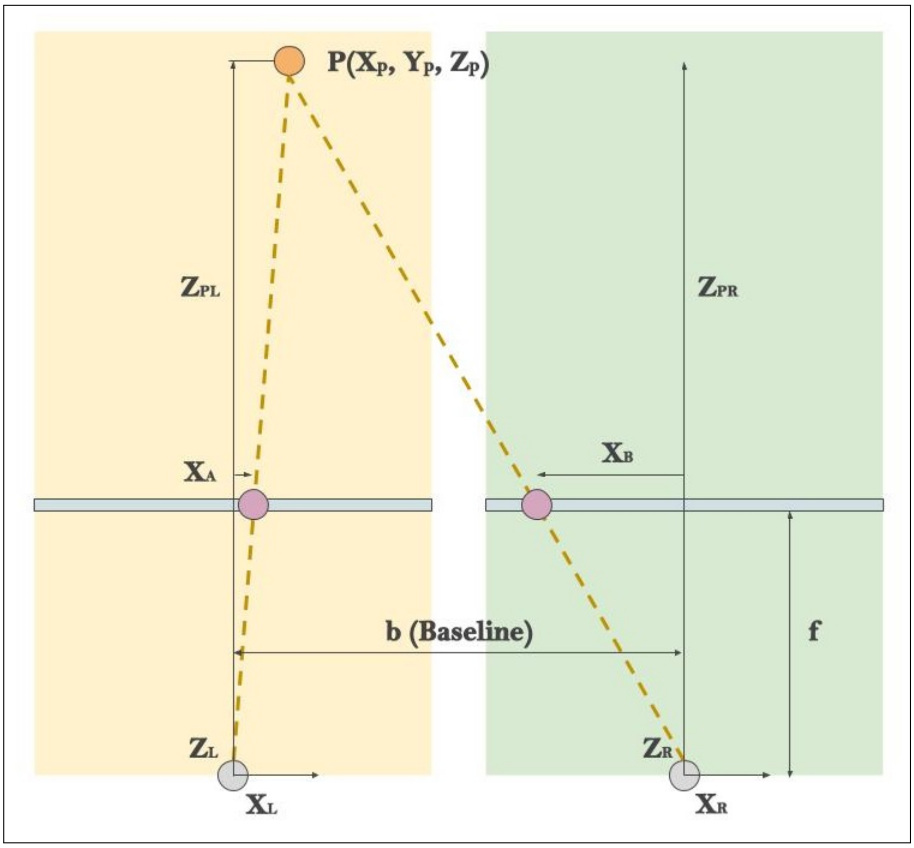


**Figure 2.16:** Use of similar triangles with basic stereo camera rig. Equations describing the relationship between disparity and distance from cameras can be found in Figure 2.15.

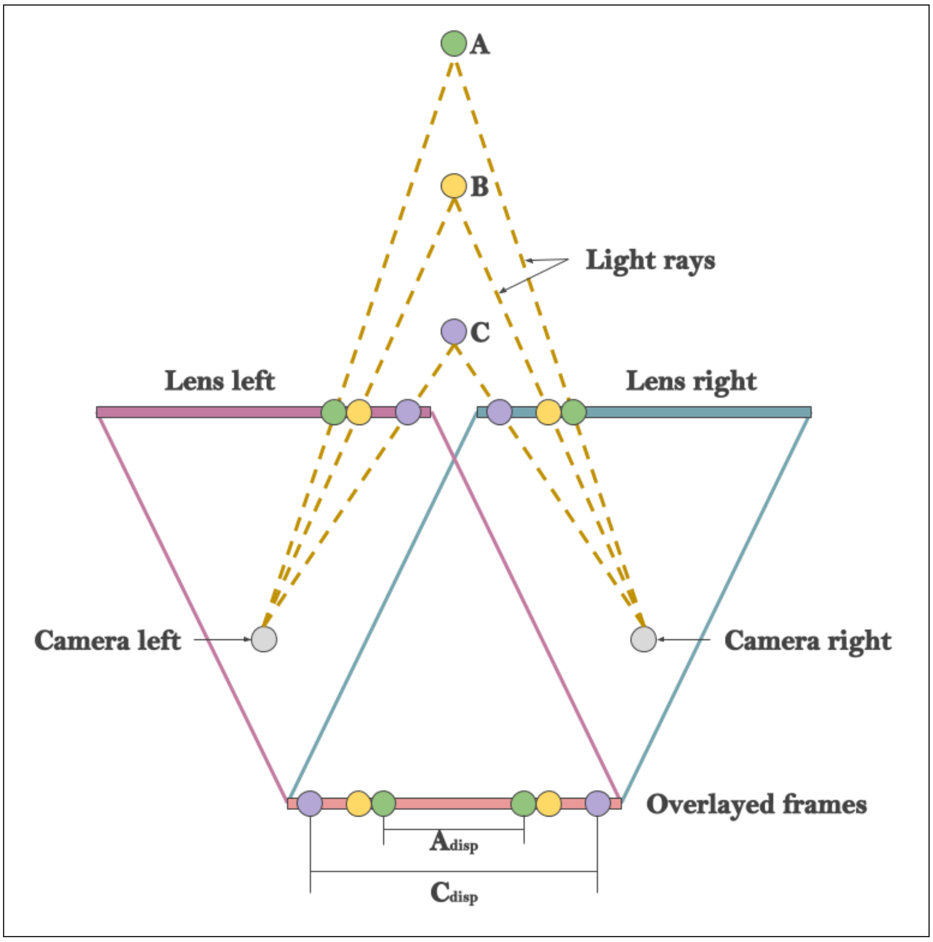


**Figure 2.17:** Disparity size with distance to object in scene.

camera lenses. Algorithms are used here to distort two frames from different viewpoints and merge them together. This requires more computation than the previously described stereo camera setup and is less accurate. The positive for this method is that it is cheap as it only requires one camera sensor.

### OBJECT RECOGNITION

Matching algorithms and the use of feature detection within multiple images to compare them are used to find and recognize objects.

Feature detection and matching algorithms refers to comparing and ranking whether a subset of an image domain matches a subset of another image domain, or a selection of other image domains. If many subsets of pixels, or features, within one image, match more than one feature in a selection of sample images, the images are correlated. Objects can be recognized by matching their features to a set of sample images with similar features and known objects. The sample image/images that best match the original image is most likely to contain the object being recognized.

A paper published by Google proposed a way to use a single computer in accurately detecting 100,000 object classes [63]. This research uses hashing algorithms similar to Google search algorithms to optimize the process of feature detection from a large sample set of images.

> *"locality-sensitive hashing to replace the dot-product kernel operator in the convolution with a fixed number of hash-table probes that effectively sample all of the filter responses in time independent of the size of the filter bank"*

Locality-sensitive hashing (LSH) is an algorithm used to search a multi-dimensional space in an efficient manner. The algorithm does not search for an exact match but an approximately close return value.

This research resulted in the development of Google's Cloud Vision platform [64].

# Chapter 3

# Design

Design thinking is the formalization of the thought that should go into a universal design. It was developed throughout the 1970s [65] [66] and thought in Stanford since the early 1990s as a fundamental way of thinking for architects and other types of designers. It is imperative that new developments are designed universally so as to increase the accessibility of the worlds population, including the visually impaired. There are five key principles to design thinking.

1. **Empathize**

   To build an effective design, the designer must know and care about the lives of the users. This involves engaging with the visually impaired community and observing the problems faced by them on a daily basis.

   This principle is addressed by conducting a focus group in conjunction with the NCBI [31]. The end users and their needs were investigated. Research into how systems have been designed to help these users in the past as also provided necessary design information.

2. **Define**

   Defining the problem being addressed in a clear and concise manner is the best way to produce a solution that will address it head on. As defined in Section 1.1, the problem that this project is addressing is:

   > *"Visually impaired people worldwide are using legacy devices for mobility and navigation which limits their quality of life."*

   Within this chapter, the author has also set out a hypothesis to prove:

   > *"Modern technologies can be used to improve the quality of life for visually impaired people."*

3. **Ideate**

   Generating a range of ideas from different perspectives and diverse outlooks allows for the best design to be concluded upon. Combining rational thoughts with imaginative thinking brings to the fore the most interesting concepts. The designer must also be open to and understand how others have addressed the problem being addressed and learn from those approaches.

   Research has been conducted into the current mainstream solutions to the problem being addressed and a literature review has been conducted.

This research has laid the foundations for how the design of a visual aid should be addressed.

4. **Prototype**

   Building a prototype is also a key step in developing an effective system design. The end user must be kept in mind when doing this. It is often the case that the prototype development phase brings forward a whole new set of problems. To combat this, it is a good approach to make this process an iterative one, developing the prototype in modular components.

   In Chapter 4 the author demonstrates how the main functionalities of the final system are explored and implemented into a prototype device. This process brought forward problems that had not been accounted for in the design stage and also provoked new opportunities that could be leveraged and incorporated into the system's design.

5. **Test**

   The final principle of design thinking involves testing. The use of testing is very important in learning about how the end users might interact with the system being designed. In Chapter 5 the prototype and design of the overall system is evaluated and discussed. This feedback loop should continuously progress the project and its effectiveness.

## 3.1 Early Design

The original design for this project involved the use of many wearable sensors on the body. The more sensors that could be utilized, in a diverse range of environments meant the more information that could be gathered about the scene. By spreading sonar and IR sensors around the body, a lot of information could be gathered about the environment that the user is in allowing decisions relating to the users mobility to be made. When today's technology is contrasted with the technology of 1995, it can be seen that a system would no longer require a robotic computer to guide the user like in [50] but the user could wear the relative sensors in a discrete manner. The sensor information could be computed and conclusions made all without the use of a bulky robotic guide.

The early design of this system as shown in Figure [3.1] contains two sonar sensors and five IR sensors. The sonar sensors are located on the upper arms of the user, directed forward. The IR sensors are located on a headband around the user's head. The information gathered from these sensors would then be processed on a smartphone which would use audio to relay the relative conclusions that have been computed back to the user.

The system prototype would have been built using an Arduino device that is connected to all sensors and wirelessly relay information to the smartphone. Upon further investigation, it was discovered that the use of IR sensors outdoors is not as effective as required. A trend was also discovered in the assistive technologies research for the visually impaired with computer vision technologies. Computer vision and the ability to compute a huge amount of information in a much shorter time period meant that the IR sensors could be replaced with a more effective technology. Upon speaking with the NCBI, it was also discovered that the majority of the information required to make effective visual aid decisions is in front of the lower half of the body. This includes stairways and curbs on pathways. Hence, the visual aid may be better suited on a lower part of the body.

The specification for this design was to be a closed system that replaced the long cane and guide dog. It was discovered that there are a lot of very good assistive technologies that could be used together to compliment each other. This resulted in a pivot for the final design that will be explored in Section 3.2 to allow the system to be extended and include new technologies as they are developed.

## 3.2 Final Design

From Table 1.2 in Section 1.2.1 it is concluded that exploiting the use of sensors or computer vision techniques at the head height is not the most efficient way of obstacle avoidance for a visually impaired person. Waist height is in a more central position to detect the fundamental obstacles that are of concern like a stairway or a curb drop-off on a pathway. This is displayed in Figure 3.2.

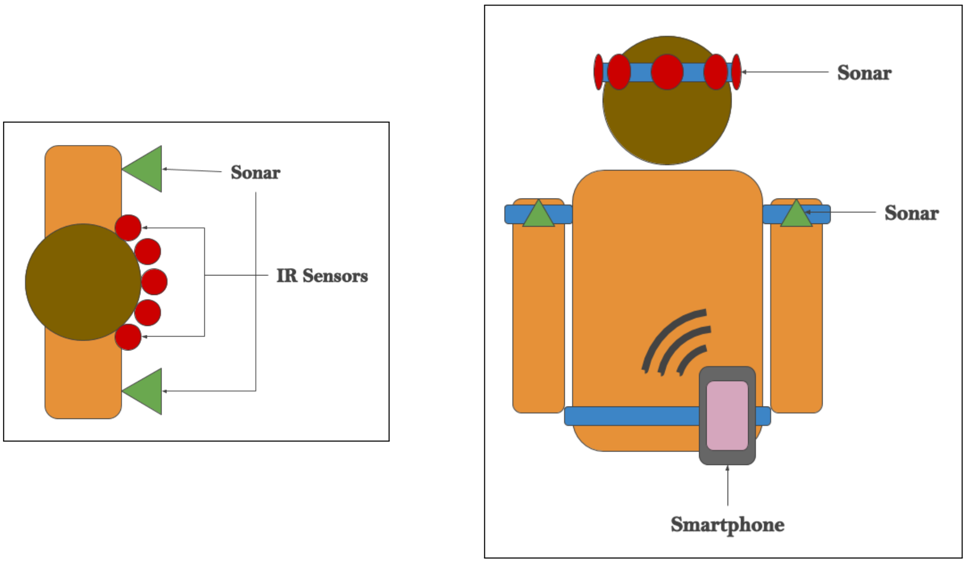


**Figure 3.1:** Early design for visual aid system using IR and sonar sensors paired with a smartphone.

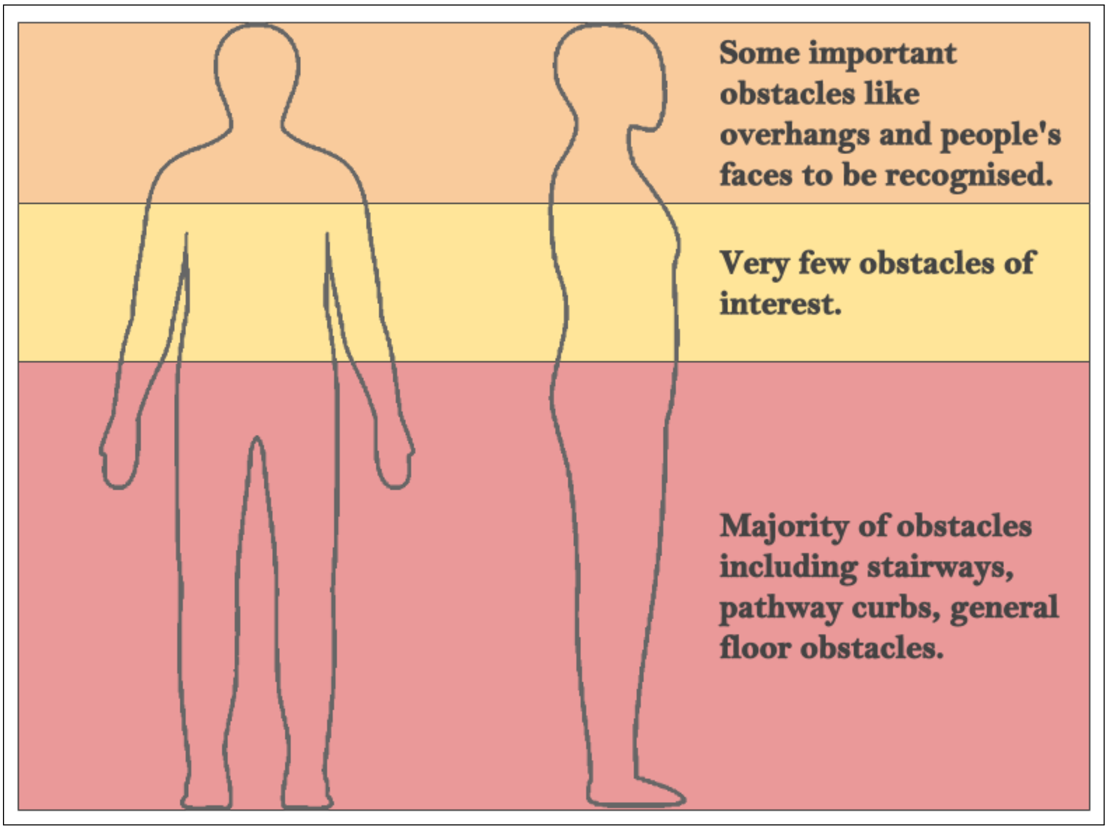


**Figure 3.2:** Obstacle locations of interest.

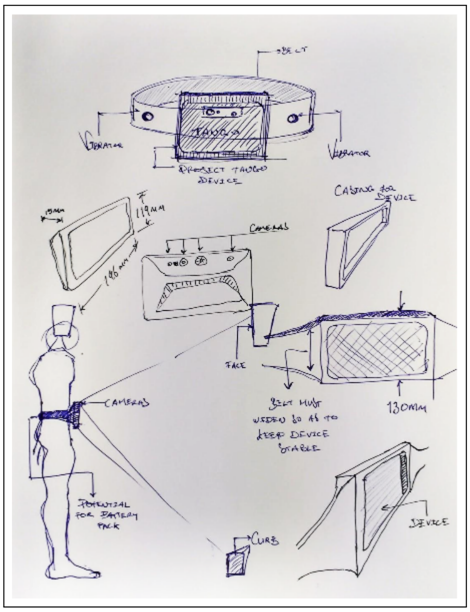


**Figure 3.3:** Belt design development.

The core technologies to be exploited in developing this system are cameras and computer vision, artificial intelligence, WiFi communication protocols and the haptic vibration within most smart devices.

This system should be an open ended and modular system. It should be possible for other software and hardware developers to use in building further features and functionalities.

The Project Tango device contains a lot of high powered sensors however the most important functionality is the ability to compute vision based tasks very fast. It uses two Movidius vision processing units to compute these functionalities in extremely short time periods. These VPUs can also be used in adding advanced artificial intelligence capabilities to the visual aid system. The fundamental development being proposed is the development of a software platform in the form of an Android application that exploits Google's Project Tango

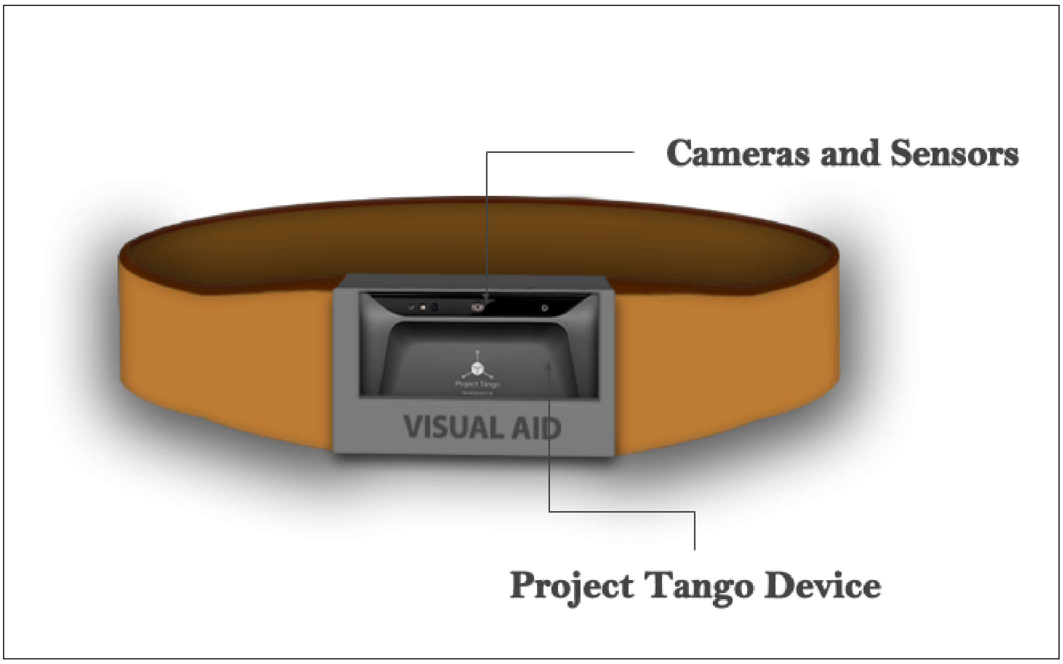


**Figure 3.4:** Basic concept for Project Tango belt.

Development Kit. Secondary to this, a belt which may have third party technologies embedded in it is also proposed. It should be possible to build modules to incorporate these third party hardwares into the Android application.

The belt seen in Figure 3.4 should be built with the main functionality of holding the Project Tango device as stable as possible to avoid motion blur and unreliable image quality for processing the camera frames.

The proposed software platform to catalyze the use of Project Tango devices as visual aids is the most important part of the design. As described in Figure 3.5, this software platform should be built using the Project Tango API. The platform should be built so as to be modular enough to allow for the easy incorporation of added features and hardwares to the system. The basic system should contain a method of haptic feedback via phone and/or smart watch as

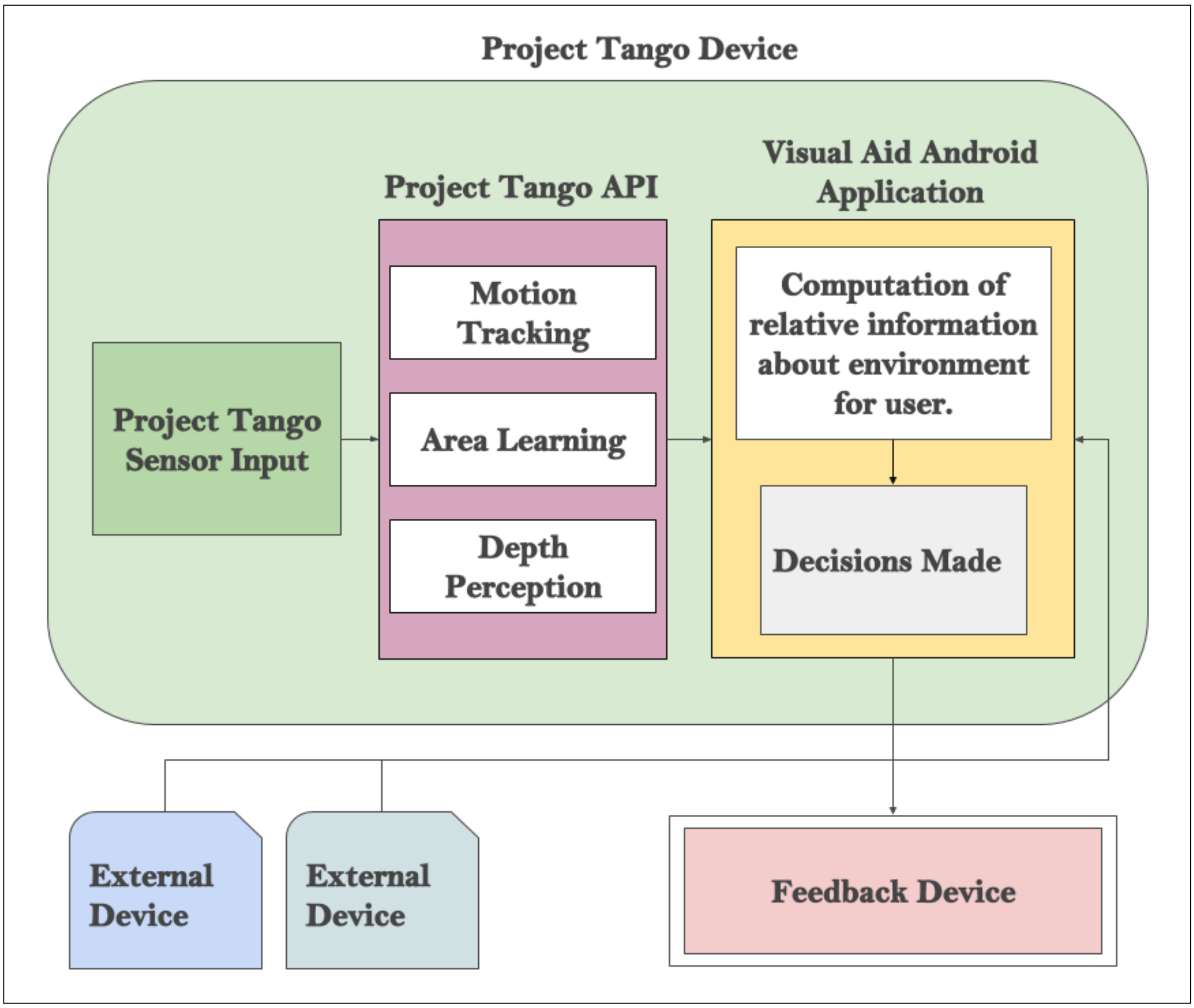


**Figure 3.5:** Basic software setup.

well as an implemented object recognition and guidance system. Figure 3.5 shows how external devices like ultrasonic and haptic sensors can be added to the system.

In addition to being a full Android phone/tablet the Project Tango device does motion tracking of the device, learns and can recognize an area and has depth perception. The Project Tango Tablet Development Kit is an Android device with a wide-angle camera, a depth sensing camera, accurate sensor timestamping, and a software stack that enables application developers to use motion tracking, area learning and depth sensing in the development of their own applications. All of these capabilities should be exploited for this visual aid software platform.

The main areas of functionality from the Project Tango device to be used in this system are:

1. Motion tracking acts as a higher resolution GPS with more reliability. Motion tracking within Project Tango devices acts similar to an optical mouse - the tracking is with reference to an origin point, this may be the entrance to a building. The device is tracked with three degrees of rotation and with regard to its forward, backwards and sideways translations. An example of the Project Tango's smooth motion tracking can be seen in Chapter 2, Figure 2.6.

   Project Tango's motion tracking works by collecting data from a fish eye camera (wide-angle fish eye cameras come close to simulating the 160 degree vision of a human eye) and inertial measurement unit (IMU). The camera images are used to identify edges and corners and tracks how much they move to gauge the distance traveled. The IMU uses an accelerometer and a gyroscope to track the acceleration and direction of the device. Camera and IMU sensor data are fused to calculate the device's movement more accurately.

In the API documentation, the position and orientation of the device is referred to as the device's pose. This data is returned from the API with two parts - A vector in meters for translation and a quaternion for the rotation of the device. The API functions of interest to request this data are:

```
TangoService_getPoseAtTime()
```

and

```
onPoseAvailable()
```

Every piece of sensor information is time-stamped accurately and each pose must be referenced with respect to a time-stamp. The camera and other sensor information that correspond to the requested pose can therefore also be requested.

Motion tracking can not be run for long periods. It must be restarted from a new origin after a lengthy period of use. This time is not defined and may change depending on the amount of device movement. With large amount of motion tracking the drift increases. This drift is addressed in the area learning functionality of the device. These limitations must be considered when using this functionality.

The most important use for this functionality in the visual aid application is to navigate around indoors. This may be helpful for finding the exit of a shop.

2. Area learning refers to the ability to recognize a space to localize the device in this space. An area description file (ADF) is built to understand the information about a space.

   This functionality is combined with motion tracking to remember areas and recognize features in them. The ADF is made up of descriptions of visual landmarks or references. Combined with motion tracking, Project Tango devices can estimate a position via localization to these visual references. It is also possible to learn areas with different levels of quality. This may be important in optimizing the visual aid functionality. An ideal environment has a variety of distinguishing statutory objects. Area learning is less accurate in dynamic and changing environments. This makes the area learning functionality of Project Tango more effective in areas like a home or office.

   Drift correction is done by closing the motion tracking loop when the device sees an area that it recognizes. This is demonstrated in Figure 3.6.

   The ADFs are saved. This has a limitation in that the ADF for an area may be different under changing lighting, this may result in different ADFs for an area at different times of day. Area learning is mostly useful for indoor tracking as there is a lot of inconsistencies with outdoor environments.

   Navigating shops, homes and offices may be greatly improved as a result

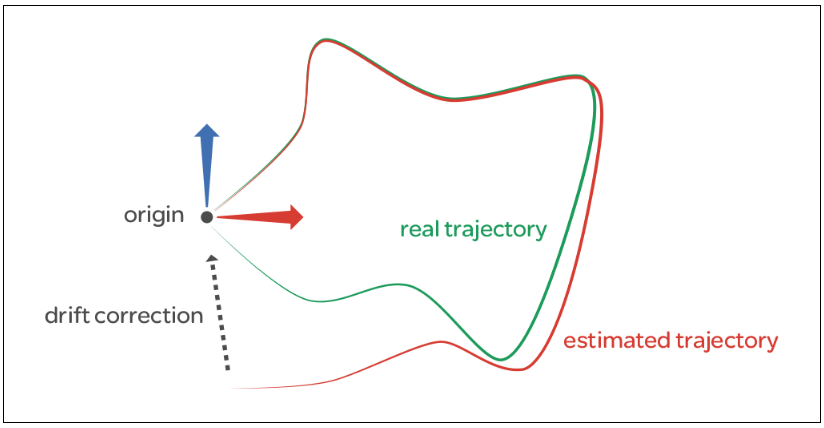


**Figure 3.6:** Drift correction in Project Tango [13].

of ADFs. A user could download the ADF for a shop or building and the visual aid application could use that to mobilize the user indoors.

3. Depth perception or point cloud capture allows for users to see distances between objects in a scene. Point cloud data is predominantly built to feed into a physics engine in a game. It can also be stored to a file for conversion to a three-dimensional model.

   This functionality can also be used to detect surfaces, obstacles or shapes in an environment. It works based off the use of three major technologies. Firstly, structured light systems that send out an IR pattern to illuminate the contours of the environment. Measuring the size of the dots in the pattern, the larger the dots get, the further away they are. Secondly, time of light sensors use IR sensors to measure how long a reflection of an infrared beam takes to return. This is similar to sonar but with light rays. These uses of IR sensors can weaken the device's ability to build thee-dimensional representations of a scene in naturally lit areas. Finally, the most important sensor used here is a stereo camera that infers depth using triangulation of a pair of camera images.

The Project Tango API allows for access to a three dimensional struct that encapsulates the point cloud info called TangoXYZij. This data structure contains information about each point in a point cloud. The i and j values correspond to the pixel in an image that has been matched to an XYZ, three-dimensional point in a point cloud. Although motion tracking and area learning may be very important interfaces to use in building extra features for a visual aid, especially for indoor use, the main concern for this system design is with the accurate point cloud information. This information can be used to extract depth data about the scene and inform the user of obstacles.

The API camera interface can be used to get the images from the camera to compute objects in a scene using:

```
TangoService_connectOnFrameAvailable(...)
```

When the images are requested, a HTTP request can be parsed for Google's Cloud Vision API to recognize up to 100,000 object types and feed the information back to the user.

As described in Figure 3.4, the visual aid Android application will take in parameters from the Project Tango API as well as other external sensors. It should then compute where the closest obstacle in the scene is located and make a request to find out what the obstacle is. The feedback mechanism should then be run to inform the user of this information.

The feedback device should be a modular module within this application design. The user has the capability to change the type of feedback that they would like. The implementation being proposed uses the haptic vibration capabilities of a smartphone as well as audio to relay information to the user, however, it should be possible to also incorporate a smartwatch and even use the smartwatch and smartphone to compliment each other in relaying the information.

To communicate the relative information to the user, the visual aid application should calculate the correct vibration sequence to run to convey the location of the closest obstacle to the user. The application should also use text to speech to inform the user of the type of obstacle that is in that location. A sequence of vibrations is being proposed to convey information to the user. This sequence is displayed in Figure 3.7. The location will be conveyed relative to its position in the captured camera images. The image is split into ”front”, ”top”, ”right” and ”left” sections. Each section will correspond to a different intensity and length of vibration. Within each of these sections is three vertical subsections. The number of vibrations will dictate which one of these subsections correspond to the image. For example, if the object is in the middle subsection of the ”right” section of the image, the vibration sequence would be made up of two vibrations with an intensity of 0.5 and a length of 500ms.

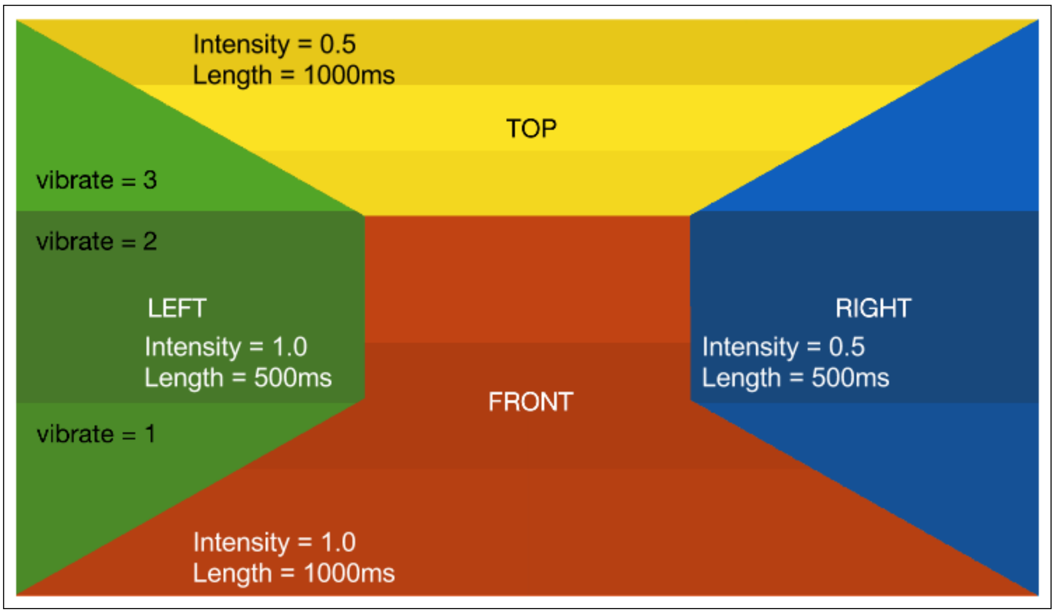


**Figure 3.7:** Vibrations sequences corresponding to image areas.

The user input will be developed using a very simple interface on the Project Tango device similar to [3] shown in Figure 2.3.

### 3.2.1 Limitations

1. Some people feel safer when others can see that they are visually impaired. Others would rather hide this impairment. It should be possible to get the belt in multiple colours. One bright colour which will inform the public that the person wearing this belt is impaired and one belt that blends subtly with current fashion trends.

2. The size of the Project Tango device is currently large as it is a tablet. Google and Lenova [67] are working towards releasing a Project Tango smartphone.

3. The device should also be secure and any communications should be safely encrypted so as to ensure that no external access can be gained to the guidance data.

4. A tablet mounted on a belt may capture distorted image frames as a result of the user's movement. The accelerometer can be used to help stabilize the captured frames.

5. Like human eyes, this device will only work with light. The capabilities of depth perception may also be weaker in sunlight as a result of the use of IR sensors.

6. The device should account for the velocity of the user. For instance, in a queue, the user may be close to the person in front of them. The device should recognize that the user's velocity is slow enough for an obstacle of close proximity to be safe. This can be done using the accelerometer.

# Chapter 4

# Implementation

The Project Tango Development Kit was not available for the development of this project. Instead of using this technology to build a prototype implementation of the design proposed in Section 3.2, a stereo camera module was developed from scratch. This method may not prove to be as powerful as Project Tango but it provides for depth-sensing test data. As well as this, Google's Cloud Vision API was incorporated and the use of haptic feedback to a smartphone using WiFi for communication was also prototyped.

The project is made up of a python server side application and Android client application. Github was used for version control [68]. Functionalities are incorporated to continually analyze an environment using a home-made stereo camera, capture a scene on manual intervals and test different aspects of the system.

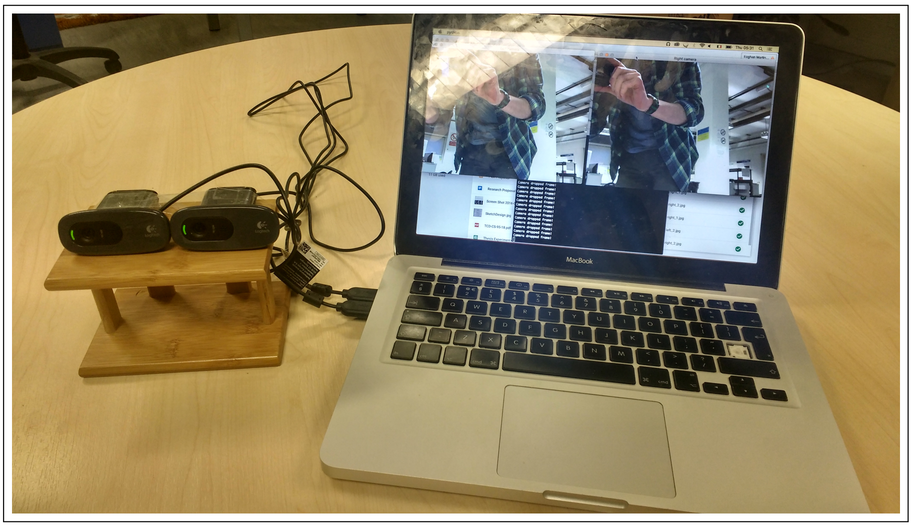


**Figure 4.1:** Home-made stereo camera setup.

## 4.1 System Overview

Two web cameras are set up to simulate a stereo camera as shown in Figure 4.1. A laptop computer is used to host a Python server which uses OpenCV to compute the stereo camera disparity map, point clouds and make requests to Google's Cloud Vision. This information is then communicated to an Android smartphone over a WiFi TCP connection.

The smartphone is used for application control and to generate a haptic feedback sequence which is used to communicate information to the user via vibrations. The main object in the scene captured by the stereo camera is simply printed to the smartphone screen. The system is shown in Figure 4.2.

The various Python application run scripts are defined in the run_scripts.py file. To run the final application with all features run:

```
sh run_scripts.py run
```

This script will wait for the smartphone client to connect and then begin pro-

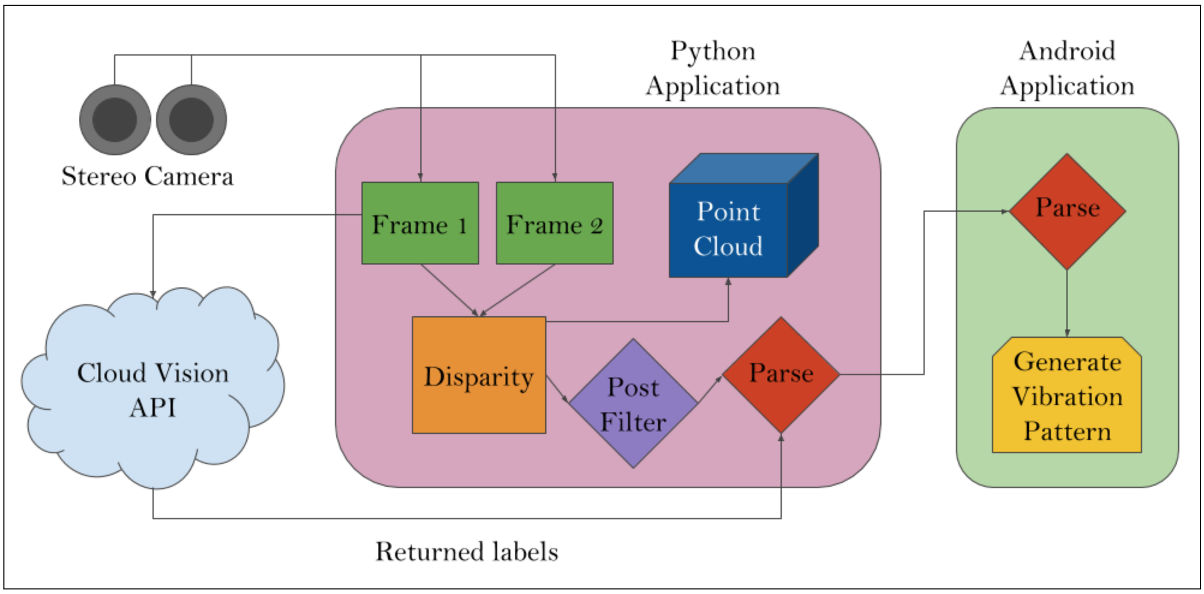


**Figure 4.2:** Implemented system layout.

cessing the camera images and communicating the computed information to the client device. The system can be paused and restarted on the client side.

## 4.2 Python Server

The Python server can be run with a variety of different parameters but overall has three fundamental functionalities. Firstly, to compute a point cloud from the stereo camera frames and calculate the area of the image which corresponds to the closest obstacle in the scene. And secondly to make a request using Google Cloud Vision to ascertain what the most prominent object in the scene is. Finally, this server communicates the data to the client.

**Figure 4.3:** Captured frames from the left and right cameras.

### 4.2.1 Open CV

Open CV is an open source computer vision library. For this project, Open CV 3.0 with Python was used on a Macbook. This library is used to import the stereo camera frames, compute the disparity image, run morphology operations and generate the point cloud .PLY file.

### 4.2.2 Depth Perception

The application imports the current frame from each web cam using OpenCV. Shown in Figure 4.3 are the left and right frames.

The disparity map for these images is then found using the settings in the settings.json file. These settings can be edited by running the sh run_scripts.sh tuner script which provides the user with a selection of sliders to edit the settings. The tuner is displayed in Figure 4.4. The parameters in this tuner contribute to the type of disparity image that is created. The function:

```
cv2.StereoSGBM_create(
  int minDisparity,
  int numDisparities,
  int blockSize,
  int P1=0, int P2=0,
  int disp12MaxDiff=0,
  int preFilterCap=0,
  int uniquenessRatio=0,
  int speckleWindowSize=0, int speckleRange=0,
  int mode=StereoSGBM::MODE_SGBM)
```

creates a StereoSGBM object. This object is used to calculate the disparity map. The StereoSGBM::compute(image[0], image[1],...) function implements Stereo-Global Block Matching (SGBM) to create a disparity map for a set of images. Stereo-Global Matching involves iteratively finding the most probable disparity for each pixel in a rectified pair of stereo images. By using block sizes instead of individual pixels, the disparities can be calculated more accurately. This is because the pixels are not perfectly horizontally aligned and therefore an epipolar line does not horizontally relate the pixels in a scene. It is possible however, to align blocks of pixels.

The parameters in settings.json are used by the StereoSGBM object to create a disparity map. They are listed in Table 4.1.

**Table 4.1** . . . continued

| Paramter | Function |
|---|---|

**Table 4.1:** Parameters used by StereoSGBM in creating a disparity map.

| **Paramter** | **Function** |
|---|---|
| minDisparity and numDisparities | Depending on the rectification accuracy of an image the minDisparity may be greater than 0. The numDisparities is the max disparity - minDisparity and corresponds to the number of depth layers in a scene. |
| blockSize | Blocksizes should be large enough so an epipolar line can be draw between both images and through a horizontal line of pixel blocks. |
| P2 and P1 | P2 should be bigger than P1. These values define the disparity smoothness. |
| preFilterCap | This is a value used in pre-processing of the images before the disparity is calculated. |
| uniquenessRatio | Percentage used to decide when a disparity value is the best for a pixel. |
| speckleRange | Maximum disparity variation within each connect component. |

The disparity image (Figure 4.5) is then used for two things:

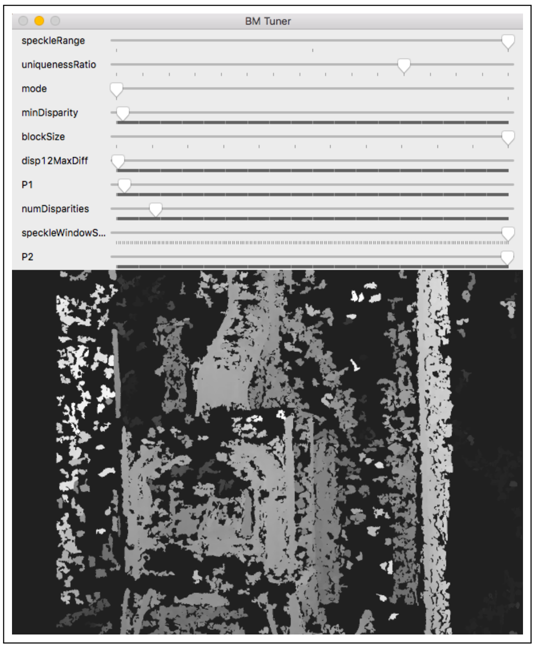


**Figure 4.4:** Settings tuner displayed by running the "sh run_scripts.sh tuner" script.

- The disparity image is manipulated using the morph_ops() function to erode, dilate, open and close the specular noise in the image and get a more accurate result. The manipulated image is then analyzed to find the areas in the scene that are closest to the cameras. These areas are returned to Android application over WiFi.

- The disparity image is also used to build a 3D point cloud map of the depth in the scene. To do this the matrix from input/disp_to_depth_mat.npy is loaded. This point cloud file is exported to the application directory as pointCloud.ply and can be opened using meshlab as shown in Figure 4.6.

After the disparity map has been constructed and the morphology operations have been run on the image, the application must determine which point

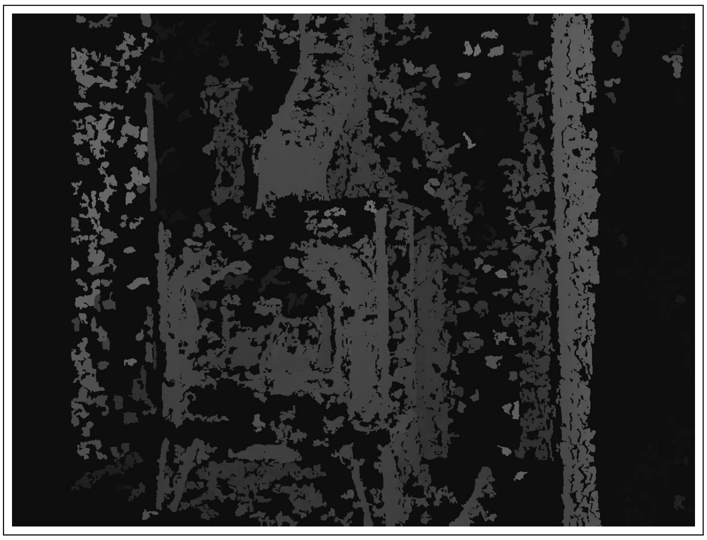


**Figure 4.5:** The disparity map produced from the camera frames in Figure 4.3.

is closest to the user. This is done by finding the area with the highest disparity value.

### 4.2.3 Point Cloud

To re-project an image to three-dimensions, a disparity to depth map matrix is used. This converts each value for disparity to a depth in the scene. The function:

```
reprojectImageTo3D(disparity, disparity_to_depth_map)
```

is used to do this. This function takes in the disparity object created using the StereoSGBM object and a disparity_to_depth_map matrix. This function adds a third-dimension to the disparity image by taking each pixel (x, y) and its corresponding disparity d=disparity(x,y) and computes the three-dimensional image as in Figure 4.7. The reconstruction of the scene in three-dimensions

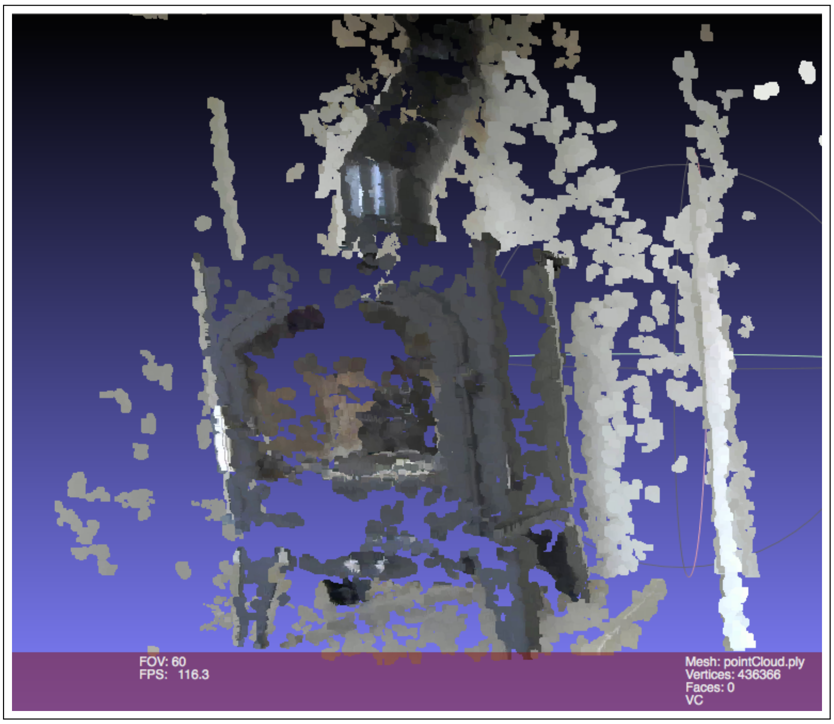


**Figure 4.6:** Point cloud creation from the images in Figure 4.3.

is used to give a developer a better idea of the accuracy of the disparity maps created. It has no application, as of yet, in assisting the end user.

$$[X\ Y\ Z\ W]^T = Q.[x\ y\ disparity(x,y)]^T$$

$$_3dImage(x,y) = (\frac{X}{W},\frac{Y}{W},\frac{Z}{W})$$

**Figure 4.7:** Computation of 3d point cloud info for pixel (x,y). Matrix Q here is the disparity_to_depth_map matrix.

### 4.2.4 Object Recognition

The captured frames are run against the Google Cloud Vision API to find objects in the scene. The request for the images in Figure 4.3 returned a result of ”Found label: wood burning stove”.

The HTTP request is built using Google Credentials and an encoded image (image_content). The JSON for the body of the request is:

```
{
  'requests': [{
    'image': {
    'content': image_content
  },
  'features': [{
    'type': 'LABEL_DETECTION',
    'maxResults': 1,
    }]
  }]
}
```

The response is also JSON and contains a label attribute which is a string containing the name of the most prominent object in the scene.

Once coordinates of the nearest point and the most prominent object in the scene have been determined, the Python server responds to the Android application with:

```
return label + ":" + str(analysis.maxLoc[0])
+ "," + str(analysis.maxLoc[1])
```

## 4.3 Android Application

The Android application has the main functionality of generating the haptic feedback sequence and displaying a string representing the most prominent object from the stereo camera images.

### 4.3.1 Communication

The Android application can be connected to the Python server by scanning a QR code for the local IP address of the machine on which the Python server application is being run. The connection is implemented using TCP over IP on a WiFi network.

The Python server is developed using the Twisted Python module. Using the Twisted module in Python it is possible to connect multiple devices to the server. This may be used for the addition of external sensors to the system.

The Android client is developed in Java and uses a separate thread to the UI thread to listen for the server responses.

Once the Android application receives a response from the server, it displays it in a list on the screen and proceeds to compute the haptic sequences for the coordinate of the closest point.

### 4.3.2 Haptic Feedback

The data received by the Android device is parsed into its seperate components. The (x,y) coordinates of the nearest point in the image is then fed into a function which generates a vibration sequence that corresponds to the respective pattern shown in Figure 3.7. It is then possible for a user to decode this sequence of vibrations to understand where the upcoming obstacle in front of a stereo camera is.

# Chapter 5

# Evaluation

This chapter evaluates the technical performance of the prototype implementation and it's ability to function as necessary. Some experiments were undertaken to investigate these aspects of the project.

**Experiment 1**

*Aim:* The investigate the speed of computation of the system.

*Methodology:* The speed of the system was tested in computing the point cloud, the most prominent object and closest point in an image pair. The speed of the system is then tested in computing the point cloud and closest point in an image pair. Followed by the computation of the most prominent object and closest point in an image pair. Finally, the system speed is tested in just computing the closest point in the image pair. Each one of these tests is run five times and the average time is calculated.

**Table 5.1:** Experiment 1 results [Tables A.1-A.4].

| **Test** | **Average Time** |
|---|---|
| Point Cloud, Closest Point and Object | 5.5 seconds |
| Point Cloud and Closest Point | 3.64 seconds |
| Object and Closest Point | 4.44 seconds |
| Closest Point | 2.12 seconds |

*Discussion:* The system is very slow when computing all three values from the image pair captured by the stereo camera. The point cloud requires the most computation as it has to generate the .PLY file. This file is only required for testing purposes. As a result, it is best to exclude this functionality in any real application of a visual aid. The object recognition takes less time to render, however, this may change depending on the user's internet speed. The computation of the closest point in the image pair is the most important part of the system. It should be returned within one second. Optimization must be addressed if the application is to be effective.

**Experiment 2**

*Aim:* The accuracy of object recognition in a scene is tested.

*Methodology:* Four common objects were tested under two different lighting conditions from four different angles, twice.

**Table 5.2:** Experiment 2 results [Tables A.5-A.12].

| **Object** | **Average Accuracy** |
|---|---|
| Chair | 68% |

**Table 5.2** . . . continued

| Object | Average Accuracy |
|---|---|
| Door | 50% |
| Object and Closest Point | 75% |
| Person | 75% |
| Table | 81.25% |

*Discussion:* Table 5.2 shows the average accuracy for recognition of a selection of objects under different light intensities. Upon closer inspection to Tables A.5-A.12 it can be seen that the object recognition is much more effective in high light intensities. An average object recognition rate of 84.375% in high light intensity environments is observed for this experiment. In low light intensities, the average accuracy rate was 53.125%. The aid will not work effectively in dim lighting. This may dictate the times of the day that it may be used safely.

**Experiment 3**

*Aim:* Investigate the accuracy of the system in determining the closest point to the user in a scene.

*Methodology:* Thirty readings were taken of environments that contained objects within a meter of the stereo camera, between one and three meters of the camera and between 3 and 5 meters of the camera. The average accuracy of the readings is recorded in Figure 5.3.

**Table 5.3:** Experiment 3 results [Table A.13].

| **Distance** | **Average Accuracy** |
|---|---|
| 0-1 meters | 80% |
| 1-3 meters | 66.6% |
| 3-5 meters | 43.33% |

*Discussion:* These results reach an overall average of **63.3%**. Project Tango's dedicated camera should achieve a higher value for this test. It can be observed that there is a significant drop-off in accuracy as the distance of obstacles increases. The current system determines a ”closest point” at all times. This means that if the closest object is more than three meters away it will be communicated to the user, with a high probability of being inaccurate. An upper limit on depth perception would allow the system to only inform the user of a close object if it is within a certain threshold, like two meters.

## 5.1 Limitations of the Evaluation

1. The system is weaker in wider spaced areas like outdoors. This could be addressed by thresholding the maximum distance that an obstacle can be from the device.

2. The system functioned poorly in low intensity lighting. To avoid bad guidance, the visual aid should not be used in the evening times. This is a drawback when compared to a long cane or guide dog which can both

function in low intensity lighting. May be addressed by the use of IR and sonar sensing when the light intensity is low.

3. The time to compute the required information is a major limitation of the system. The computation takes a minimum of two seconds using a Macbook with a 1.2GHz processor and low resolution (420x640 pixel) cameras. To apply this system to a real test case it must be optimized.
4. The objects in the scene are not be converted to speech. They are displayed as text on the screen.
5. Using WiFi is not a consistent method of communication for a mobile visual aid system as the user may leave the network. This could be replaced using a Bluetooth pairing connection between the server and client.
6. A laptop is not potable. The server should be moved to a Project Tango device which can be mounted to a waist belt. This device could do the computer vision computation and communicate over Bluetooth with a smartphone and/or smart watch.

# Chapter 6

# Conclusion

This dissertation outlines a system to assist the visually impaired in mobility. The proposed system design demonstrates how Google's Project Tango device can be used to develop a technology based visual aid. An implementation of a visual aid using a home-made stereo camera and Android device is also demonstrated and evaluated.

Both the proposed design and implemented visual aid system incorporate the use of computer vision techniques. Literature and research in this field has also been explored within this report. The key technology to developing a technology based visual aid is computer vision.

It can be concluded that the implemented visual aid system is not a feasible solution for mobility to a person with a visual impairment. It can, however, be deduced from the evaluation of this system that with more optimized algorithms and dedicated hardware, a stereo camera based computer vision system could greatly improve the mobility of a visually impaired person, thus improving their quality of life.

## 6.1 Future Work

Implementation of a working prototype has been successful. Focus should now be focused on improving the prototype communication and optimization with the ultimate goal of migrating the system to a Project Tango device. Firstly, the WiFi communication protocols can be extended to be implemented over Bluetooth. This is a much more reliable and network independent method of communication.

Next, text to speech functionality should be implemented. This is important for converting the objects in the scene from text to speech for the user to understand.

After communication and text to speech functionality have been implemented, the system can be migrated to a Project Tango device. These devices are optimized for computer vision tasks and have an embedded stereo camera. Once a portable and high speed prototype has been developed, the system can be feasibility tested with a focus group of visually impaired people.

The barrier to entry is high for assistive technologies like this visual aid as it is required to be accurate to be effective. A qualitative system evaluation involving real visually impaired people to reveal underlining system qualities and characteristics would be an important step in bringing the solution closer to market.

# Appendix A

# Experiment Results

**Table A.1:** Experiment 1: Point Cloud, Closest Point and Object.

| **Run** | **Time (seconds)** |
|---|---|
| 1 | 5.6 |
| 2 | 5.4 |
| 3 | 5.8 |
| 4 | 5.3 |
| 5 | 5.4 |
| Average | **5.5** |

**Table A.2:** Experiment 1: Object and Closest Point.

| **Run** | **Time (seconds)** |
|---|---|
| 1 | 3.8 |
| 2 | 3.6 |
| 3 | 3.6 |
| 4 | 3.7 |
| 5 | 3.5 |

**Table A.2** . . . continued

| Run | Time (seconds) |
|---|---|
| Average | **3.64** |

**Table A.3:** Experiment 1: Closest Point and Object.

| **Run** | **Time (seconds)** |
|---|---|
| 1 | 4.3 |
| 2 | 4.6 |
| 3 | 4.5 |
| 4 | 4.4 |
| 5 | 4.4 |
| Average | **4.44** |

**Table A.4:** Experiment 1: Closest Point.

| **Run** | **Time (seconds)** |
|---|---|
| 1 | 2.2 |
| 2 | 2.1 |
| 3 | 2.2 |
| 4 | 2.0 |
| 5 | 2.1 |
| Average | **2.12** |

**Table A.5:** Experiment 2: Recognition of a chair (first run).

| **Chair** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | True Positive | True Positive | True Positive |
| Low Light Intensity | False Negative | True Positive | False Negative | True Positive |

**Table A.6:** Experiment 2: Recognition of a chair (second run).

| **Chair** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | False Negative | True Positive | True Positive | True Positive |
| Low Light Intensity | False Negative | True Positive | False Negative | True Positive |

**Table A.7:** Experiment 2: Recognition of a door (first run).

| **Door** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | False Negative | True Positive | True Positive |
| Low Light Intensity | False Negative | False Negative | False Negative | True Positive |

**Table A.8:** Experiment 2: Recognition of a door (second run).

| **Door** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | True Positive | False Negative | True Positive |
| Low Light Intensity | True Positive | False Negative | False Negative | True Positive |

**Table A.9:** Experiment 2: Recognition of a person (first run).

| **Person** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | True Positive | False Negative | True Positive |
| Low Light Intensity | True Positive | False Negative | True Positive | True Positive |

**Table A.10:** Experiment 2: Recognition of a person (second run).

| **Person** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | True Positive | True Positive | True Positive |
| Low Light Intensity | False Negative | True Positive | True Positive | True Positive |

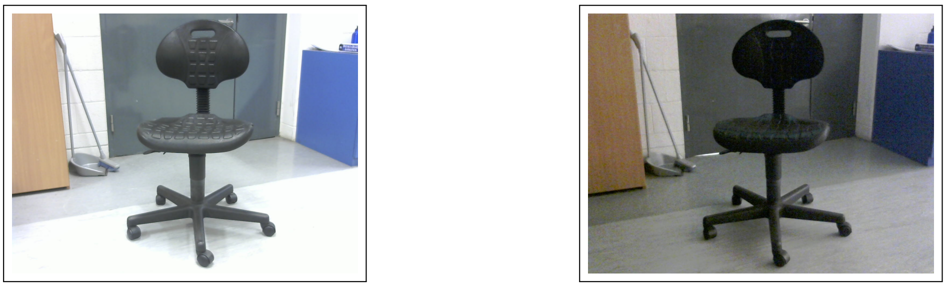


**Figure A.1:** High and low light intensity levels for experiment 2.

**Table A.11:** Experiment 2: Recognition of a table (first run).

| **Table** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | True Positive | False Negative | True Positive |
| Low Light Intensity | False Negative | True Positive | True Positive | True Positive |

**Table A.12:** Experiment 2: Recognition of a table (second run).

| **Table** | **Angle 1** | **Angle 2** | **Angle 3** | **Angle 4** |
|---|---|---|---|---|
| High Light Intensity | True Positive | True Positive | True Positive | True Positive |
| Low Light Intensity | True Positive | True Positive | False Negative | True Positive |

For Experiment 2, the overall average True Positive result for high light intensity object recognition is 84.375%. The overall average True Positive result for low light intensity object recognition is 53.125%.

**Table A.13:** Experiment 3: Accuracy of closest point determination.

| **Reading** | **0-1 meter** | **1-3 meters** | **3-5 meters** |
|---|---|---|---|
| 1 | TRUE | TRUE | FALSE |
| 2 | TRUE | TRUE | FALSE |
| 3 | FALSE | TRUE | TRUE |
| 4 | TRUE | TRUE | TRUE |

**Table A.13** . . . continued

| Reading | 0-1 meter | 1-3 meters | 3-5 meters |
|---|---|---|---|
| 5 | TRUE | FALSE | FALSE |
| 6 | FALSE | TRUE | FALSE |
| 7 | TRUE | TRUE | TRUE |
| 8 | TRUE | TRUE | FALSE |
| 9 | TRUE | TRUE | FALSE |
| 10 | TRUE | FALSE | TRUE |
| 11 | TRUE | TRUE | FALSE |
| 12 | TRUE | TRUE | TRUE |
| 13 | FALSE | FALSE | FALSE |
| 14 | TRUE | TRUE | TRUE |
| 15 | TRUE | TRUE | FALSE |
| 16 | TRUE | FALSE | FALSE |
| 17 | TRUE | TRUE | FALSE |
| 18 | TRUE | TRUE | FALSE |
| 19 | FALSE | FALSE | TRUE |
| 20 | TRUE | TRUE | TRUE |
| 21 | FALSE | FALSE | TRUE |
| 22 | TRUE | TRUE | FALSE |
| 23 | TRUE | TRUE | TRUE |
| 24 | TRUE | TRUE | FALSE |
| 25 | TRUE | FALSE | FALSE |
| 26 | TRUE | TRUE | TRUE |
| 27 | FALSE | FALSE | FALSE |
| 28 | TRUE | FALSE | TRUE |
| 29 | TRUE | FALSE | FALSE |
| 30 | TRUE | TRUE | TRUE |

**Table A.13** . . . continued

| Reading | 0-1 meter | 1-3 meters | 3-5 meters |
|---|---|---|---|
| Accuracy | **80%** | **66.6%** | **43.3%** |

# Bibliography

[1] Smartcane image. [Online]. Available: smartcane.saksham.org/overview

[2] Orcam. [Online]. Available: orcam.com

[3] Project ray. [Online]. Available: project-ray.com

[4] Aftershokz: Headphones that deliver unrivaled situational awareness and comfort. [Online]. Available: aftershokz.com

[5] Horus technology. [Online]. Available: horus.tech/en

[6] Project tango motion tracking image. [Online]. Available: developers.google.com/project-tango/overview/motion-tracking#common_use_cases

[7] M. H. Kamaludin, “Sonar assistive device for visually impaired people,” *Universiti Teknologi Malaysia*, 2015. [Online]. Available: researchgate.net/publication/282464822_Sonar_Assistive_Device_for_Visually_Impaired_People

[8] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, “Fingerio: Using

active sonar for fine-grained finger tracking," *ACM CHI*, 2016. [Online]. Available: fingerio.cs.washington.edu/fingerio.pdf

[9] J. Tanz. Kinect hackers are changing the future of robotics. [Online]. Available: wired.com/2011/06/mf_kinect

[10] M. Brock, "Supporting blind navigation using depth sensing and sonification," *2013 ACM conference on Pervasive and ubiquitous computing*, 2013. [Online]. Available: ubicomp.org/ubicomp2013/adjunct/adjunct/p255.pdf

[11] Brock, "A navigation system for the visually impaired: A fusion of vision and depth sensor," *2013 ACM conference on Pervasive and ubiquitous computing*, 2013. [Online]. Available: hindawi.com/journals/abb/2015/479857

[12] D. Scharstein, R. Szeliski, and H. Hirschmller. Stereo vision research resource. [Online]. Available: vision.middlebury.edu/stereo/

[13] Google area learning. [Online]. Available: developers.google.com/project-tango/overview/area-learning#improving_the_trajectory

[14] W. H. O. (2014) Icidh-2: International classification of functioning and disability. [Online]. Available: who.int/mediacentre/factsheets/fs282/en/

[15] M. Hersh and M. A Johnson, *Assistive Technology for Visually Impaired and Blind People*, first edition ed. London: Springer-Verlag, 2008, ch.

Disability and Assistive Technology Systems, pp. 2–4. [Online]. Available: web.iitd.ac.in/~pvmrao/assistive_book.pdf

[16] W. H. O, “Icidh-2: International classification of functioning and disability,” *Beta-2 draft, Full Version*, p. 191, 1999. [Online]. Available: sustainable-design.ie/arch/Beta2full.pdf

[17] C. Barnes, *Disabled People in Britain and Discrimination: A Case for Anti-Discrimination Legislation*, first edition ed. Hurst & Company, 1991.

[18] R. Velzquez, *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization*, first edition ed. Springer, 2010, ch. Wearable Assistive Devices for the Blind.

[19] M. Hersh and M. A Johnson, *Assistive Technology for Visually Impaired and Blind People*, first edition ed. London: Springer-Verlag, 2008, ch. Disability and Assistive Technology Systems, p. 29. [Online]. Available: web.iitd.ac.in/~pvmrao/assistive_book.pdf

[20] Assistive technology industry association (atia). [Online]. Available: atia.org/

[21] Assistive technology conference of new england. [Online]. Available: assistivetechnologyconference.com

[22] Atec conference. [Online]. Available: ateconference.com

[23] Aaate conference 2017. [Online]. Available: aaate2017.eu

[24] Global elevate awards. [Online]. Available: globalelevate.com

[25] Global elevate awards. [Online]. Available: resna.org/about/awards

[26] B. C. C, "Disabled and elderly assistive technologies," 2015. [Online]. Available: bccresearch.com

[27] P. W. C, "The wearable future," October, 2014. [Online]. Available: pwc.com/us/en/technology/publications/assets/pwc-wearable-tech-design-oct-8th.pdf

[28] I. D. C, "U.s. wearable computing device 20142018 forecast and analysis," 2014.

[29] N. Hunn, "The market for smart wearable technology - a consumer centric approach," March, 2015. [Online]. Available: nickhunn.com/wp-content/uploads/downloads/2014/08/The-Market-for-Smart-Wearables.pdf

[30] N. C. B. I, "National mobility report: Mobility experiences and perceptions of blind and vision impaired persons," 2012. [Online]. Available: ncbi.ie/sites/default/files/exec_summary_web.pdf

[31] *Meeting with NCBI Member and 6 Friends of the Organisation (May 2016)*, ser. NCBI Meeting, NCBI, Clondalkin, Dublin, 2016.

[32] Knfb reader. [Online]. Available: knfbreader.com/

[33] Smartcane. [Online]. Available: smartcane.saksham.org

[34] Ultracane. [Online]. Available: ultracane.com

[35] Blindsquare. [Online]. Available: blindsquare.com

[36] Transition technologies s.a. [Online]. Available: seeingassistant.tt.com.pl/move

[37] Cm project spotlight: Sentiri. [Online]. Available: chaoticmoon.com/chaos-theory/cm-project-spotlight-sentiri

[38] Wayfindr. [Online]. Available: wayfindr.net

[39] B. B. C. Tube app: Guide for blind people on london underground. [Online]. Available: bbc.com/news/health-31757593

[40] Cities unlocked. [Online]. Available: citiesunlocked.org.uk

[41] Lechal. [Online]. Available: lechal.com

[42] Google project tango. [Online]. Available: google.com/atap/project-tango/

[43] D. Cox and T. Dean, “Neural networks and neuroscience-inspired computer vision,” *Current Biology*, vol. 24, pp. 921–929, 2014. [Online]. Available: cell.com/current-biology/abstract/S0960-9822(14)01039-2

[44] Argus ii retinal prosthesis system. [Online]. Available: secondsight.com/g-the-argus-ii-prosthesis-system-pf-en.html

[45] Pixium vision homepage. [Online]. Available: pixium-vision.com/en

[46] A. Jason, “Intra-ocular device,” Patent 14/522 710, 10 24, 2014.

[47] B. Otis and B. Parviz. Introducing our smart contact lens project. [Online]. Available: googleblog.blogspot.ie/2014/01/introducing-our-smart-contact-lens.html

[48] J. A. Brabyn, “New developments in mobility and orientation aids for the blind,” *IEEE Transactions on Biomedical Engineering*, 1982. [Online]. Available: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.159&rep=rep1&type=pdf

[49] M. P. B, “An experimental system for auditory image representations,” *IEEE Transactions on Biomedical Engineering*, 1992.

[50] G. Lacey, K. M. Dawson-Howe, and D. Vernon, “Personal adaptive mobility aid for the infirm and elderly blind,” *Technical Report TR-CS-95-18*, 1995.

[51] A. D’Amico and R. Pittenger, “A brief history of active sonar,” *Aquatic Mammals Journal*, p. 426, 2009.

[52] P. S. L. , “Depth mapping using projected patterns,” Patent US 12/522,171, 5 13, 2010.

[53] J. C. Lee, “Hacking the nintendo wii remote,” *IEEE Intelligent Systems*, 2008.

[54] L. Porzi, S. Messelodi, C. M. Modena, and E. Ricci, “A smart watch-based gesture recognition system for assisting peo-

ple with visual impairments," *VENTURI (FP7-288238)*, 2011. [Online]. Available: venturi.fbk.eu/wp-content/uploads/2011/10/A-Smart-Watch-based-Gesture-Recognition-System-for-Assisting-People-with-Visua pdf

[55] P. R. Sanketi and J. M. Coughlan, "Anti-blur feedback for visually impaired users of smartphone cameras," *The Smith-Kettlewell Eye Research Institute*, 2010.

[56] A. Bhowmick, S. Prakash, R. Bhagat, V. Prasad, and S. M. Hazarika, "Intellinavi : Navigation for blindbased on kinect and machine learning," *Multi-disciplinary Trends in Artificial Intelligence by Springer*, 2014. [Online]. Available: researchgate.net/publication/271328979_IntelliNavi_Navigation_for_Blind_Based_on_Kinect_and_Machine_Learning

[57] Raytrix homepage. [Online]. Available: raytrix.de/

[58] Lytro homepage. [Online]. Available: lytro.com

[59] R. Ng, "Digital light field photography," 2006. [Online]. Available: stanford.edu/class/ee367/reading/Ren%20Ng-thesis%20Lytro.pdf

[60] M. Levoy and P. Hanrahan, "Light field rendering," *SIGGRAPH*, 1996.

[61] D. Lanman, "Mask-based light field capture and display," *Degree of Doctor of Philosophy in the School of Engineering at Brown*

*University*, 2011. [Online]. Available: alumni.media.mit.edu/~dlanman/thesis/Lanman_PhD_thesis.pdf

[62] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," *CVPR*, vol. 1, pp. 519–526, 2006. [Online]. Available: vision.middlebury.edu/mview/seitz_mview_cvpr06.pdf

[63] T. Dean, M. A. Ruzon, M. Segal, J. Shlens, S. Vijayanarasimhan, and J. Yagnik, "Fast, accurate detection of 100,000 object classes on a single machine," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2013, 2013. [Online]. Available: static.googleusercontent.com/media/research.google.com/en//pubs/archive/40814.pdf

[64] Google cloud vision api. [Online]. Available: cloud.google.com/vision/

[65] H. Simon, *The Sciences of the Artificial*, first edition ed. Cambridge: MIT Press, 1969.

[66] R. McKim, *Experiences in Visual Thinking*, first edition ed. Brooks/Cole Publishing Co., 1973.

[67] Lenovo: The world's first project tango-powered smartphone. [Online]. Available: lenovo.com/projecttango

[68] E. Martin. Visual aid. [Online]. Available: github.com/eoghanmartin/VisualAid